

Tilburg University

Improving Garch Volatility Forecasts

Klaassen, F.J.G.M.

Publication date:
1998

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):

Klaassen, F. J. G. M. (1998). *Improving Garch Volatility Forecasts*. (CentER Discussion Paper; Vol. 1998-52). Econometrics.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Improving GARCH Volatility Forecasts

by Franc Klaassen *

Department of Econometrics, Tilburg University

May 13, 1998

Abstract

Many researchers use GARCH models to generate volatility forecasts. We show, however, that such forecasts are too variable. To correct for this, we extend the GARCH model by distinguishing two regimes with different volatility levels. GARCH effects are allowed within each regime, so that our model generalizes existing regime-switching models that allow for ARCH terms only. The empirical application on U.S. dollar exchange rates shows that our model indeed yields better volatility forecasts than single-regime GARCH and that the allowance for GARCH terms besides ARCH terms can be crucial for the forecast quality.

Key words: GARCH, regime-switching, volatility, forecasting, exchange rates.

JEL classification: C52, C53, F31.

*Correspondence to: Franc Klaassen, Department of Econometrics, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, the Netherlands; tel: +31-13-4668229; fax: +31-13-4663280; E-mail: F.J.G.M.Klaassen@kub.nl.

I would like to thank Frank de Jong, Harry Huizinga, Theo Nijman, Casper de Vries and Bertrand Melenberg, as well as Arthur van Soest and Geert Bekaert for their helpful comments.

1 Introduction

Exchange rates play a key role in the international economy. There is a lot of currency trading on financial markets, and also in international goods trade at least one of the trading partners has to deal with a foreign currency. Not only the levels of exchange rates are important, also their volatility is. For example, exchange rate volatility is important in the pricing of currency options and may also affect international trade.

It is well-known that volatility is time-varying in high-frequency data and that periods of high volatility tend to cluster. To capture this, many authors have used autoregressive conditional heteroskedasticity (ARCH) models, as introduced by Engle (1982) and extended to generalized ARCH (GARCH) in Bollerslev (1986).¹ Such models usually improve the fit a lot compared with a constant variance model and, as Andersen and Bollerslev (1997) claimed, GARCH models provide good volatility forecasts.

In this paper we show that GARCH forecasts are, nevertheless, too variable. We try to improve on the GARCH forecasts by extending the GARCH model such that this weakness disappears. Therefore, our model contains an extra source of volatility persistence by distinguishing two regimes with different volatility levels. GARCH formulas are used to capture any volatility persistence within the regimes. The resulting regime-switching GARCH model indeed yields better volatility forecasts, both one-period and multi-period-ahead, in-sample as well as out-of-sample.

What is the reason behind the high variability of standard GARCH volatility forecasts? It is well-known that GARCH models often imply a high volatility persistence of individual shocks. For example, for their stock return data Hamilton and Susmel (1994) showed that a shock this week will have a nonnegligible effect on the variance a full year later. Therefore, GARCH models are unable to capture the fact that shocks sometimes have the effect of "relieving pressure" on the system, that is, a shock is followed by a period of low instead of high volatility. The source of volatility persistence that we add to standard GARCH models makes that individual shocks can, but need not be persistent. Our empirical results make clear that this greater flexibility in the persistence of shocks removes the too high a variability of the volatility forecasts. Apparently, it is the high volatility persistence that makes standard GARCH forecasts too variable.

Lamoureux and Lastrapes (1990), among others, showed that the high volatility persistence in GARCH models may originate from structural changes in the variance

¹See Bollerslev, Chou and Kroner (1992) for an overview of the GARCH literature.

process. They demonstrated that any shift in the unconditional variance is likely to lead to misestimation of the GARCH parameters in such a way that they imply too high a volatility persistence. For example, if the variance is high but constant for some time and low but constant otherwise, persistence of such high and low homoskedastic periods already results in volatility persistence. A GARCH model, which cannot capture persistence of such periods, puts all volatility persistence in the persistence of individual shocks.

Structural changes in the variance process can originate from changes in economic policy, such as the change in U.S. monetary policy in 1979, when the Federal Reserve reduced the money growth rate to bring down inflation and to stop the dollar's fall. Sudden shifts may also result from more exogenous changes in the economic environment, such as the OPEC oil crises.

One possibility to allow for periods with different unconditional variances is, of course, by introducing deterministic shifts into the variance process, but this is rather ad hoc. A popular approach to endogenize changes in the data generating process is the use of a Markov regime-switching model. Hamilton (1989) introduced this model to describe the U.S. business cycle, which is characterized by periodic shifts from recessions to booms and vice versa. In our context of exchange rate volatility, a Markov process can be used to govern the switches between regimes with different variances. Even a simple regime-switching model with constant regime-specific variances already captures a lot of the volatility persistence, as our empirical results show.

To capture the remaining conditional heteroskedasticity, the Markov approach can be combined with ARCH models, as Cai (1994) and Hamilton and Susmel (1994) showed. Their regime-switching ARCH models have two ways to capture volatility persistence, namely the persistence of regimes and the ARCH effects within a regime. We show empirically that for some series a moderate number of ARCH terms is indeed sufficient to capture all conditional heteroskedasticity.

A potential drawback of regime-switching ARCH models is that only ARCH models are allowed within a regime, not GARCH models. Our empirical results show that this is not only a theoretical disadvantage, but that it is also important from a practical point of view, as some series are characterized by long persistence of shocks within a regime. One GARCH term can capture such long persistence much more parsimoniously than a large number of ARCH terms. Moreover, as we show, neglecting the long persistence can result in even worse volatility forecasts than those generated by a single-regime GARCH model. For that reason Gray (1996a) modified the approaches of Cai (1994) and Hamilton and Susmel (1994) such that also GARCH effects are allowed in each

variance regime.

Our variance specification also allows for GARCH effects. The crucial difference with Gray's model is the way the unobserved regime indicator is integrated out. We use all available information, whereas Gray uses only part of it. The important result of this is that we get much simpler, first-order recursive variance forecasting formulas. Together with the first-order recursive structure of the likelihood function, this makes the model very useful from a practical point of view.

Another feature of our model is that volatility persistence of individual shocks is allowed to be different across regimes. Our empirical results show that in periods of high volatility shocks are more persistent than in low volatility periods. Cai (1994) and Hamilton and Susmel (1994) did not account for this, but Gray (1996a) did. A shift from the high variance to the low variance regime is thus only part of the way to explain the "pressure relieving" effect discussed above. The other part results from the fact that the regime shift also implies a shift to the low-persistence regime. Besides the introduction of two variance regimes, this leads to additional flexibility regarding the persistence of shocks. Recall that it is the spuriously high persistence in single-regime GARCH models that causes the high variability of the forecasts.

In the next section we will formally introduce our regime-switching GARCH model. The recursive estimation process is described in section 3, after which section 4 describes alternative ways to determine the regime the process was likely in at a particular time. Section 5 presents the recursive forecasting formula that makes volatility forecasting in our regime-switching GARCH model much easier than in Gray's (1996a) model. In section 6 we describe the twenty years of daily U.S. dollar exchange rates vis-a-vis the British pound, the German mark and the Japanese yen that we use in our empirical application. The empirical results are also in this section. Finally, section 7 concludes.

2 Regime-Switching GARCH Model

In this section we present regime-switching models to capture the conditional heteroskedasticity that is present in many high-frequency financial time-series. We first discuss regime-switching models in the literature so far, and then, having showed the problems involved in them, we introduce our regime-switching GARCH model.

Let S_t denote the logarithm of a spot exchange rate at time t . Since there is substantial empirical evidence that exchange rates themselves are not covariance stationary, but that their first differences are (see Hakkio and Rush (1989)), we analyze the return $s_t = 100(S_t - S_{t-1})$. Thus, s_t is the percentage depreciation from time $t-1$ to time t . Let I_{t-1} denote the information set at $t-1$, consisting of all returns up to time $t-1$.

As the paper is mainly focussed on modeling return volatility rather than its level, we assume for simplicity that the **conditional mean** of s_t is constant:

$$s_t = \mu + \varepsilon_t, \quad (1)$$

where the error term ε_t satisfies $E_{t-1}\{\varepsilon_t\} = E\{\varepsilon_t | I_{t-1}\} = 0$.² Thus, $\mu = E_{t-1}\{s_t\}$ is the expected depreciation.

To model the **conditional variance** of s_t many authors have used GARCH models. This improves the fit a lot, and the implied persistence of individual shocks in the variance is usually very high (see Hamilton and Susmel (1994)). This high volatility persistence may be spurious and originate from instability of the GARCH parameters, as Lamoureux and Lastrapes (1990) demonstrated. As argued in the introduction, we take account of this instability by the introduction of two variance regimes as a second source of volatility persistence. Let r_t be the (unobserved) regime at time $t-1$, where the first regime is identified as the low variance one. We will now discuss four specifications for $V_{t-1}\{\varepsilon_t | r_t\}$, the variance conditional on observed information I_{t-1} and on the unobserved regime being r_t .

One possible specification for the conditional variance is a GARCH(1,1) model. In a regime-switching context, this would define

$$V_{t-1}\{\varepsilon_t | r_t\} = \omega_{r_t} + \alpha_{r_t} \varepsilon_{t-1}^2 + \beta_{r_t} V_{t-2}\{\varepsilon_{t-1} | r_{t-1}\}, \quad (2)$$

where the current regime only determines the parameters, that is, the intercept ω_{r_t} , the ARCH parameter α_{r_t} and the GARCH parameter β_{r_t} . However, the previous conditional variance $V_{t-2}\{\varepsilon_{t-1} | r_{t-1}\}$ cannot be computed, since we do not observe regimes. Moreover, $V_{t-2}\{\varepsilon_{t-1} | r_{t-1}\}$ depends on $V_{t-3}\{\varepsilon_{t-2} | r_{t-2}\}$, which depends on $V_{t-4}\{\varepsilon_{t-3} | r_{t-3}\}$, and so on. Consequently, the conditional variance in (2) depends on the entire sequence of variance regimes up to time t . Each different combination of previous regimes thus leads to a different conditional variance at time t . Since the number of possible combinations grows exponentially with t , this leads to an enormous number of regime paths to t . This renders estimation intractable, since the likelihood contribution of observation t is constructed by integrating out over all possible regime paths.

In the literature so far we find two different approaches to circumvent this problem of path dependence. First, Cai (1994) and Hamilton and Susmel (1994) essentially removed the GARCH term, which is the cause of the path dependence, from (2) and

²It is possible to incorporate, for example, autoregressive terms in the conditional mean without making the formulas that follow essentially different.

thus used only ARCH terms.³ Therefore, $V_{t-1}\{\varepsilon_t | r_t\}$ in (2) only depends on the current regime r_t .⁴

However, as Gray (1996a) argued, the problem of path dependence can be solved without giving up the potentially important persistence effects of a GARCH term. Gray first integrated out the unobserved regime r_{t-1} in $V_{t-2}\{\varepsilon_{t-1} | r_{t-1}\}$ in (2) using the information observable at time $t-2$ and then used the aggregated variance in a GARCH formula:

$$V_{t-1}\{\varepsilon_t | r_t\} = \omega_{r_t} + \alpha_{r_t} \varepsilon_{t-1}^2 + \beta_{r_t} E_{t-2} \{V_{t-2}\{\varepsilon_{t-1} | r_{t-1}\}\}. \quad (3)$$

By integrating out the regime r_{t-1} he obtained a conditional variance that only depends on the current variance regime r_t , not on previous regimes. He thus avoided the problem of path dependence by a slight redefinition of the GARCH(1,1) formula in (2). He showed that the likelihood function can be computed recursively, similar to that in a single-regime GARCH model.

There is one important problem with Gray's method: generating variance forecasts $V_{t-1}\{s_{t-1+l}\}$ for some lead $l > 1$ turns out to be very complicated. The approach we will use does not have this problem with forecasting. Furthermore, our approach also allows for persistence effects through a GARCH term, without suffering from the problem of path dependence.

The improvement of our model with respect to forecasting is due to two crucial differences with Gray's model. First, as the expectation in (3) shows, Gray integrated out the regime r_{t-1} at time $t-2$. We postpone this till $t-1$, the time at which we want to compute the conditional variance. This allows us to use more observable information when integrating out the previous regime. The extra data embodies information about previous regimes and is thus useful.

The second crucial difference is that, when integrating out the regime r_{t-1} , Gray did not use the information that the variance regime at time t is in the conditioning information of $V_{t-1}\{\varepsilon_t | r_t\}$. Particularly if regimes are highly persistent, the variance regime at time t will give much information about the regime at $t-1$. In contrast to Gray, we do use that information.

³In both papers the authors use more than one lag in the ARCH specification. All results in the present paper can be generalized to include also more than one lag in a straightforward manner. It is only for the sake of exposition that we restrict the number of lags to one.

⁴Cai (1994) and Hamilton and Susmel (1994) used slightly different models in which $V_{t-1}\{\varepsilon_t | r_t\}$ not only depends on the current but also on a few recent regimes. The essential point is that the conditional variance depends on a small number of regimes, so that the short regime paths can be integrated out quite easily.

In formula, our regime-switching GARCH(1,1) model is described by

$$V_{t-1}\{\varepsilon_t | r_t\} = \omega_{r_t} + \alpha_{r_t} \varepsilon_{t-1}^2 + \beta_{r_t} E_{t-1} \{ V_{t-2}\{\varepsilon_{t-1} | r_{t-1}\} | r_t \}. \quad (4)$$

Thus, we take the mean of (2) over the previous regime r_{t-1} conditional on all available information, that is, the observable information at time $t-1$, I_{t-1} , and the information about the current variance regime r_t . By construction, $V_{t-1}\{\varepsilon_t | r_t\}$ only depends on the current variance regime r_t , so there is no path dependence. In section 5 we shall show that this approach indeed leads to relatively simple, recursive formulas for the conditional variance $V_{t-1}\{s_{t-1+l}\}$ of future exchange rate returns.

To ensure positivity of $V_{t-1}\{\varepsilon_t | r_t\}$ for all t , we assume $\omega_{r_t} > 0$ and $\alpha_{r_t}, \beta_{r_t} \geq 0$. We also assume that the "unconditional" variance $V\{\varepsilon_t | r_t\}$ exists for both regimes for all ω_1 and ω_2 . Necessary conditions for this are in appendix A, which also provides a formula for the unconditional variance. The necessary conditions are similar to the necessary (and sufficient) condition $\alpha + \beta < 1$ in the one-regime GARCH(1,1) model.

It is clear from above that, even within a regime, we allow for volatility persistence. The exact nature of this persistence is allowed to differ between regimes. For instance, if shocks are more persistent in periods of high volatility than in periods of low volatility, the regime specific parameters in (4) are able to capture this asymmetry. This has important consequences for capturing the "pressure relieving" effect of some large shocks. That is, sometimes a shock relieves the market from tensions, so that the period following the shock is quiet rather than volatile. Any regime-switching model can capture this to some extent by a shift from the high volatility to the low volatility regime. However, our regime-switching model with different parameters across regimes has an additional source of neglecting large recent shocks. If the low variance regime is the short persistence regime, the large shock will be out of the market very soon after the switch to the low variance regime. This extra flexibility regarding the volatility persistence of shocks will prove important in the empirical analysis. In this respect, our model generalizes the models in Hamilton and Susmel (1994) and Cai (1994), even if GARCH terms are not present. After all, their regime variances only differ by a multiplicative or additive constant, respectively, not by differences in the ARCH parameters.

Given our specification of the conditional mean and variance, the regime-specific **conditional return distribution** is completed by the assumption that, conditional on the observable information I_{t-1} and on the regime being r_t , the error term has a t -distribution with ν_{r_t} degrees of freedom, zero mean and variance $V_{t-1}\{\varepsilon_t | r_t\}$:

$$\varepsilon_t | r_t, I_{t-1} \sim t(\nu_{r_t}, 0, V_{t-1}\{\varepsilon_t | r_t\}). \quad (5)$$

We choose a t-distribution instead of a normal one to allow for leptokurtosis even in the regime-specific conditional distribution. We will show empirically that this is particularly important in regime-switching models, since it improves the stability of the regimes. In case of normality, a large innovation in the low volatility period will lead to a switch to the high volatility regime earlier, even if it is a single outlier in an otherwise quiet period. Note that the t-distribution includes the normal distribution as a limiting case as the degrees of freedom go to infinity.

So far, we have described the exchange rate return process conditional on the regime. To complete the model, we now describe the **regime process**. As in Hamilton (1989), we assume that the latent variance regime indicator r_t follows a Markov process with constant staying probabilities

$$p_{t-1}(r_t | r_{t-1}) = p(r_t | r_{t-1}) = \begin{cases} p_{11} & \text{if } r_t = r_{t-1} = 1 \\ p_{22} & \text{if } r_t = r_{t-1} = 2. \end{cases} \quad (6)$$

Thus, the conditional probability that the current regime is the low or high variance regime depends on the past only through the most recent regime r_{t-1} . Both staying probabilities are strictly between zero and one, so that there is no absorbing regime.

Equations (1), (4), (5) and (6) describe the complete regime-switching GARCH model. The first three equations determine the density of the exchange rate s_t conditional on I_{t-1} and on the variance regime r_t , as we will show in the next section. The last equation describes the regime process, which will be needed to integrate out the unobserved regimes. Note that the standard, one-regime GARCH(1,1) model is a special case of our model, with, of course, the regime process left out.

3 Estimation

We estimate the regime-switching GARCH model by maximum likelihood. To obtain the likelihood function, we first need the density of the exchange rate return at time t conditional on only observable information. Let $p_{t-1}(s_t)$ denote this density evaluated at an exchange rate return equal to s_t . We have

$$p_{t-1}(s_t) = \sum_{r_t=1,2} p_{t-1}(s_t | r_t) \cdot p_{t-1}(r_t). \quad (7)$$

The first term on the right-hand-side, $p_{t-1}(s_t | r_t)$, denotes the density of the exchange rate return at time t evaluated at the value s_t conditional on I_{t-1} and on the regime having value r_t . This density follows from equations (1), (4) and (5). These imply that, conditional on the data through time $t-1$ and on the regime r_t , exchange rate returns

have the following t-distribution:

$$s_t | I_{t-1}, r_t \sim t(\nu_{r_t}, \mu, V_{t-1}\{\varepsilon_t | r_t\}), \quad (8)$$

where

$$V_{t-1}\{\varepsilon_t | r_t\} = \omega_{r_t} + \alpha_{r_t}(s_{t-1} - \mu)^2 + \beta_{r_t} E_{t-1}\{V_{t-2}\{\varepsilon_{t-1} | r_{t-1}\} | r_t\}. \quad (9)$$

Since the regime in the conditioning information in (8) is unobservable, it has to be integrated out using the second term in (7), $p_{t-1}(r_t)$, which is the conditional probability that the variance regime at t has value r_t . This makes the density of s_t conditional on only observable information a mixture of two t-distributions, with a time-varying mixing parameter. Appendix B presents a simple formula to compute $p_{t-1}(r_t)$. It also gives a simple formula for the probability $p_{t-1}(r_{t-1} | r_t)$, which is needed to compute the expectation in (9). Furthermore, both probabilities and all other regime probabilities in the paper can be derived from only one key probability, which is $p_{t-1}(r_{t-1})$. Using similar techniques as in Gray (1996a), we show in appendix B that this key probability has a first-order recursive structure, which simplifies its computation a lot.

Summarizing the computation of the density $p_{t-1}(s_t)$, we have shown that one only needs the first-order recursive key probability for the two regime outcomes, the previous return s_{t-1} and two previous variances $V_{t-2}\{\varepsilon_{t-1} | r_{t-1}\}$. Therefore, computing $p_{t-1}(s_t)$ for all t is a first-order recursive process, similar to that of a one-regime GARCH(1,1) model. This is a very important implication of the path independent modeling of the conditional error variance in (4). Integrating out the previous regime at each recursion implies that one can forget about earlier regimes, and this makes the calculation of the densities quite fast.

These densities can now be used to build the sample log-likelihood $\sum_{t=1}^T \log(p_{t-1}(s_t))$ with which all parameters in the regime-switching GARCH model can be estimated. To start up the recursive computation of the log-likelihood, we set the required variables equal to their expectation without conditioning on the information set. In appendix A we derive the "unconditional" variance $V\{\varepsilon_t | r_t\}$, which is used to start-up $V_{t-1}\{\varepsilon_t | r_t\}$ at $t=1$.

4 Regime Inference

An interesting aspect of regime-switching models is that one can endogenously determine when the variable under consideration was likely in the high or low variance regime. The regime probabilities used so far, such as $p_{t-1}(r_t)$ and $p_{t-1}(r_{t-1})$, use only information available at time $t-1$. In the literature, $p_{t-1}(r_t)$ is called an *ex ante* regime

probability, while $p_{t-1}(r_{t-1})$ is a filter probability (see Gray (1996a)). However, if one is interested in determining ex post when the process was likely in a particular regime at time t , one can use all information there is at the time $\tau \geq t$ one wants to compute the probability. Let $p\tau(r_t)$ denote the ex post probability.

An ex post regime probability that is of special interest is the smoothed probability $p_T(r_t)$, which uses all information in the dataset ($\tau = T$). It gives the most efficient answer to the question which regime the process was likely in at time t . Smoothed regime probabilities have been used in many papers for such regime inference (see Gray (1996a) and Cai (1994)).

In appendix C we show that all ex post regime probabilities $p\tau(r_t)$ can be computed recursively. The algorithm is based on Gray (1996b). It links the ex ante probabilities, which are used during estimation, directly to the ex post probabilities by iterating forward from the ex ante to the ex post probabilities. Note that for $\tau = t$ this algorithm gives the filter probabilities.

5 Forecasting

One of the advantages of our regime-switching GARCH model over Gray's (1996a) model is the ease of multi-period-ahead forecasting. Suppose one is interested in the variance of the l -period-ahead exchange rate return s_{t-1+l} given information available at time $t-1$. For notational convenience, let τ denote the future time, that is, $\tau = t-1+l \geq t$. The variance of interest is

$$V_{t-1}\{s\tau\} = \sum_{r\tau=1,2} p_{t-1}(r\tau) \cdot V_{t-1}\{\varepsilon\tau | r\tau\}. \quad (10)$$

An important implication of our way of modeling the conditional variance in (4) is that $V_{t-1}\{\varepsilon\tau | r\tau\}$ can be computed in a recursive manner using a formula analogous to the one Engle and Bollerslev (1986) derived for the standard, one-regime GARCH model. Starting from $V_{t-1}\{\varepsilon_t | r_t\}$, which is a by-product of the calculation of the density $p_{t-1}(s_t)$, one can compute $V_{t-1}\{\varepsilon\tau | r\tau\}$ for $\tau > t$ by iterating forward on

$$V_{t-1}\{\varepsilon_{t+i} | r_{t+i}\} = \omega_{r_{t+i}} + (\alpha_{r_{t+i}} + \beta_{r_{t+i}}) \cdot E_{t-1} \{V_{t-1}\{\varepsilon_{t+i-1} | r_{t+i-1}\} | r_{t+i}\}, \quad (11)$$

for $i=1, \dots, \tau-t$. This recursive formula is proved in appendix D.⁵ It makes the computation of $V_{t-1}\{s\tau\}$ in (10) quite easy. These multi-period-ahead volatility forecasts will be compared to the standard, one-regime GARCH(1,1) forecasts in the empirical application of the next section.

⁵Expressions for the regime probabilities involved in (10) and (11) are in appendix B.

6 Empirical Results

In this section we examine empirically the quality of the regime-switching GARCH model developed in section 2. First, we describe the data. Then we estimate the model and analyze the differences between the regime-switching GARCH model and the standard, single-regime GARCH(1,1) model. We also compare our regime-switching GARCH model with regime-switching ARCH models similar to the ones used by Cai (1994) and Hamilton and Susmel (1994), so that we can examine whether the GARCH effects, which they neglect, are practically relevant. Finally, we examine whether the introduction of the regime-switches to the GARCH model has indeed resulted in better volatility forecasts.

6.1 Data

We use three major U.S. dollar exchange rates, namely, the dollar vis-a-vis the British pound, the German mark and the Japanese yen.⁶ We have 4982 daily observations for the exchange rate return s_t from January 3, 1978 to July 23, 1997. All rates have been obtained from Datastream.

Panel A of figures 1, 2 and 3 at the end of section 6.2 gives an indication of the volatility clustering of the three exchange rates under consideration over the sample period. We present the squared returns, s_t^2 , instead of the returns themselves to get a clearer distinction between periods of high and low volatility. This is also useful when assessing the quality of the volatility forecasts, which will be done below. All three squared return plots show substantial volatility clustering, indicating the potential usefulness of allowing for conditional heteroskedasticity.

The plots also demonstrate that shocks sometimes have a long effect on subsequent volatility, but that shocks can also be followed by a period of low volatility. To show whether single-regime GARCH, regime-switching ARCH and regime-switching GARCH models can capture this, let us consider figure 1A as an example. The large peak in the squared return plot for the British pound on March 27, 1985 was followed by about half a year of substantial volatility. Single and multi-regime ARCH models cannot capture such long-run persistence of individual shocks. However, GARCH models can, and one typically finds a high sum of the ARCH and GARCH parameters in the standard, one-regime GARCH models, indicating high volatility persistence of individual shocks (see Hamilton and Susmel (1994)).

Figure 1A also shows that shocks sometimes have the effect of "relieving pressure"

⁶U.S. dollar exchange rates are defined as the dollar price of one unit of foreign currency.

on the system, so that a shock is followed by a period of low instead of high volatility. For instance, the G-5 Plaza announcement on September 22, 1985 to bring about a dollar depreciation had a sharp effect on the dollar the next day, as the second largest peak in figure 1A makes clear. The sharp fall on that day apparently relieved the foreign exchange market from the tensions that had resulted from the sharp dollar appreciation in the years before, as the foreign exchange market was relatively quiet in the remaining part of 1985. This feature cannot be explained by the large persistence of individual shocks that is typically implied by a standard, one-regime GARCH model. It can, however, be captured by regime-switching models, since these allow for a switch from a high to a low volatility regime in such a situation. Our regime-switching model with GARCH effects can thus capture both the "pressure relieving" effect and the large volatility persistence of shocks, as shown in the previous paragraph.

In table 1 we report some descriptive statistics of three exchange rate returns. The second part of this table analyzes the autocorrelation in the returns and their squares. Starting with the squared returns, the significant first-order autocorrelations point at conditional heteroskedasticity for all three series.⁷ Estimates for higher-order autocorrelations are not reported separately, but are combined in Box-Pierce type statistics Q_{10}^s ; they also indicate conditional heteroskedasticity.

As the autocorrelations of the squared returns suggest heteroskedasticity, we use heteroskedasticity-consistent standard errors for the autocorrelation statistics in the returns themselves. Only the British pound shows significant autocorrelation, which is completely due to the first-order coefficient. For that reason, we use a first-order autoregressive term in the mean equation (1) for the British pound.

6.2 Estimation Results

This subsection presents the estimation results for the regime-switching GARCH model. For comparison, we also estimate four other models. These benchmark models are two single-regime models, namely the constant variance model and the GARCH(1,1) model, and two regime-switching ARCH models, namely the ARCH(0;0) and ARCH(4;4) model. We use the notation ARCH(Q_1 ; Q_2) for a regime-switching model with Q_1 ARCH terms in the first regime and Q_2 in the second regime. The ARCH(4;4) model is comparable with those of Cai (1994) and Hamilton and Susmel (1994). It is, however, somewhat more general, as the ARCH parameters are allowed to differ freely across regimes, whereas in Cai (1994) and Hamilton and Susmel (1994) the parameters only differ by an additive or multiplicative constant, respectively.

⁷We always use a significance level of 5%.

For all models we use a t-distribution instead of a normal one. In the regime-switching models we do this to make the persistence of regimes less sensitive to outliers, as will be explained below. For uniformity, we also use t-distributions in the single-regime models. The usefulness of t-distributed errors in the single-regime GARCH model has been shown in earlier studies, such as Engle and Bollerslev (1986).

Tables 2, 3 and 4 present the maximum likelihood results for the British pound, German mark and Japanese yen, respectively. For easier comparison of the models, we present the "unconditional" variances $\sigma_1^2 = V\{\varepsilon_t \mid r_t = 1\}$ and $\sigma_2^2 = V\{\varepsilon_t \mid r_t = 2\}$ instead of the intercepts ω_1 and ω_2 in the conditional variance formula (4). They can be computed with the formulas in appendix A. Moreover, we present the inverse of the degrees of freedom of the t-distribution; testing for conditional normality then boils down to testing whether ν^{-1} differs significantly from zero.

Single-regime GARCH

As is typically found, the standard, one-regime GARCH(1,1) model provides a much better fit than the constant variance model. For instance, the increase in log-likelihood of the GARCH model over the constant variance model is 244.34 for the British pound, so that ARCH and GARCH effects are statistically very important. Furthermore, the sum of the ARCH and GARCH parameters ($\alpha + \beta$) is large for all three series pointing at high volatility persistence of individual shocks. This has also been found in earlier papers, for instance West and Cho (1995). Panel D of figures 1, 2 and 3 shows the estimated variance series $\hat{V}_{t-1}\{s_t\}$ for the three series. The volatility persistence appears from the gradual decrease of the conditional variance after a shock.

The high volatility persistence of shocks in the single-regime GARCH model may well indicate parameter instability, as we showed before. We estimate regime-switching models to analyze whether the high volatility persistence is indeed spurious.

Regime-switching ARCH(0;0)

Let us first consider the regime-switching ARCH(0;0) model in which persistence of regimes is the only source of volatility clustering. Tables 2, 3 and 4 show that all three regime-switching models clearly distinguish between a low and a high volatility regime, where the unconditional variance in the latter is about three times as large.

The variance regimes are also persistent, since the staying probabilities p_{11} and p_{22} are all above 0.975. To get a better idea about the amount of persistence that such staying probabilities imply, we compute the expected duration of the high variance regime. Conditional on being in this regime ($r_t = 2$), its expected duration is (see

Hamilton (1989))

$$\sum_{l=1}^{\infty} l \cdot P\{r_t=2, \dots, r_{t+l-1}=2, r_{t+l}=1 | r_t=2\} = \sum_{l=1}^{\infty} l \cdot (p_{22})^{l-1}(1-p_{22}) = (1-p_{22})^{-1}.$$

For a typical ARCH(0;0) staying probability of 0.98, this implies an expected duration of 50 (working) days, which is about 2.5 months.

The introduction of variance regimes captures much of the volatility persistence in the data. To show this, we analyze the normalized residuals. Since in our model the variance depends on an unobserved regime, normalizing the residuals is not as easy as usual. One first has to integrate out the regime, as in (10). The square root of the resulting variance can then be used as the normalizing factor.

Tables 2, 3 and 4 present tests for heteroskedasticity in the normalized residuals. The first-order autocorrelations ρ_1^s and the aggregate autocorrelation tests Q_{10}^s for the squared normalized residuals show that the conditional heteroskedasticity in the normalized residuals has greatly reduced when going from the constant variance model to the regime-switching model with constant regime-specific variances. However, the heteroskedasticity tests also make clear that there is still heteroskedasticity left. Apparently, there is also volatility clustering within a regime.

Regime-switching ARCH(4;4)

Cai (1994) and Hamilton and Susmel (1994) tried to capture the volatility clustering within regimes by ARCH dynamics. The heteroskedasticity tests for the yen show the usefulness of this approach. A regime-switching ARCH(4;4) model is sufficient to capture all conditional heteroskedasticity in the dollar-yen exchange rate returns, although at the cost of a number of extra parameters.

Allowing for only ARCH effects in a regime-switching model is, however, insufficient for the two European currencies, as the aggregate autocorrelation tests Q_{10}^s show. This remaining conditional heteroskedasticity can be attributed to the high variance regime, as the ARCH(4;4) estimates for the high variance regime point at potentially longer persistence in the high variance regime only; for the low variance regime they show that even less than four ARCH terms would have been enough. Note that this also illustrates the importance of letting the ARCH parameters differ across regimes. This is in contrast with the models Cai (1994) and Hamilton and Susmel (1994) used, since they restricted the variances in both regimes to be the same apart from an additive of multiplicative scaling parameter, respectively.

An event that appeared to have a particularly long effect on the dollar-pound volatility is the crash of the Exchange Rate Mechanism of the European Monetary System in

September, 1992 (see figure 1A). Taking only four lags is probably insufficient to generate good conditional variance estimates for this period. The GARCH(1,1) estimated variance in figure 1D seems to capture the gradual decrease in volatility better than the ARCH(4;4) variance in figure 1G.

Regime-switching GARCH

The long persistence in the high volatility regime for the pound and the mark, as indicated by the ARCH(4;4) results, show the potential usefulness for GARCH terms as a parsimonious representation of the volatility persistence within regimes. Indeed, the sum of ARCH and GARCH parameters in the high volatility regime is large for the pound and the mark, and there is no heteroskedasticity left in the normalized residuals. Also the log-likelihood increases a lot after the introduction of GARCH terms in the high volatility regime: 42.15 for the pound and 22.71 for the mark. This shows that the persistence of individual shocks is largest in the dollar-pound exchange rate. For the yen this persistence is smallest, as the log-likelihood increase is -1.91.

Besides the volatility persistence within regimes, the regime-switching GARCH model also uses the regime-persistence as a source of volatility persistence. The persistence of regimes is illustrated by the plots of the estimated smoothed probability, $p_T(r_t)$, that the process is in the high variance regime (see panel B of figures 1, 2 and 3). We see that the two European currencies have experienced much less regime shifts than the Japanese yen. Apparently, sudden shifts in the variance are more important for the description of the yen than of the European currencies, where the conditional variance is more governed by smooth transitions (GARCH effects) from high volatility periods to low ones. This is also clear from the increase in the log-likelihood when introducing regime-specific GARCH effects to the ARCH(0;0) model, which has only regime-shifts to capture conditional heteroskedasticity. For the yen the increase is only 7.70, whereas for the pound it is 70.29 and for the mark 33.62.

An issue closely related to the persistence of regimes is the allowance for extra leptokurtosis by a t-distribution within a regime. Without this, the persistence of the, for example, low volatility regime would have been lower, since then a large sudden change in the exchange rate would have been considered earlier as a shift to the high volatility regime. This is illustrated by figure 1F, which gives the smoothed regime probabilities of the regime-switching GARCH model for the British pound under the restriction of normality. We see that under normality more regime-switches occur.

Comparing single-regime and multi-regime GARCH

So far, we have shown that regime-switching GARCH models are capable of capturing all volatility clustering, whereas regime-switching ARCH models may fail. But standard, one-regime GARCH models also seem to capture the volatility clustering, at least according to the autocorrelation statistics in tables 2, 3 and 4. What is then the reason for introducing an extra source of volatility persistence by allowing for two regimes? Figure 1H shows the answer. This plot illustrates the essential difference between single-regime and regime-switching GARCH models. It contains the conditional variance estimates of both GARCH models for the British pound for 1985 and 1986 only. The long effect of the largest shock in the data (March 27, 1985) on subsequent volatility appears to be captured well by both models. However, the sharp fall in the dollar one day after the G-5 Plaza announcement on September 22, 1985 is not dealt with correctly by the one-regime GARCH model. It overestimates the volatility after this event, because of the large volatility persistence of individual shocks. The regime-switching model, however, is able to capture this "pressure relieving" effect, and will thus lead to better volatility forecasts in such periods. It can cut off the effect of large shocks in two ways. First, a switch from the high volatility to the low volatility regime leads to a sharp decrease in the variance. This regime-switch is also apparent from figure 1E, which plots the ex ante regime probability $p_{t-1}(r_{t-1})$ that the process is in the high volatility regime. Secondly, after the regime-switch shocks have a much shorter effect, as the low volatility regime is also the low persistence regime. This second channel is only present if one does not restrict the GARCH parameters in both regimes to be equal. Given the large parameter differences, this channel appears to be an important way to forget large recent shocks.

In summary, our regime-switching GARCH model takes account of two significant aspects of exchange rate distributions, namely the occurrence of many large changes and the clustering of them. The first aspect is already captured in our constant variance model by the t-distribution for the innovations. This characteristic is also present in the four other models. For the second aspect, the clustering of large changes, the regime-switching GARCH model distinguishes two sources. The first one is the persistence of periods with different unconditional variances. This source is absent in the two single-regime models. Secondly, our model allows for volatility persistence even within regimes (in contrast with ARCH(0;0)) and allows for long-run persistence, which is in contrast with ARCH(4;4). In the next subsection we will analyze whether these model differences affect the forecast quality.

6.3 Forecasting Performance

In this subsection we compare both the in-sample and out-of-sample forecasts generated by the five models discussed above. The forecasts are computed for two horizons, namely the one-day horizon, which corresponds to the data frequency, and the ten-days horizon.

In-sample forecasting

The in-sample forecasts at time $t-1$ of the variance of the return at some future time τ , $\hat{V}_{t-1}\{s\tau\}$, follow from (10) after substitution of the estimation results of subsection 6.2. Since the conditional variance is the conditional expectation of $(s\tau - \mu)^2$, we compare $\hat{V}_{t-1}\{s\tau\}$ with $(s\tau - \hat{\mu})^2$. We first take the **one-day-ahead forecasts**, so that $\tau = t$.

Following many other papers, such as Gray (1996a), the first forecast statistic we consider is the root mean squared error (RMSE), defined as the square root of $\frac{1}{T} \sum_{t=1}^T \left((s\tau - \hat{\mu})^2 - \hat{V}_{t-1}\{s\tau\} \right)^2$. From panel A of tables 5, 6 and 7 we see that the regime-switching GARCH model outperforms the standard, one-regime GARCH model for all series.⁸ Apparently, the introduction of regimes into a GARCH model not only leads to a better fit, but also to better forecasts.

We also see that for the British pound, the series with the highest volatility persistence of shocks, the outperformance of regime-switching GARCH over single-regime GARCH is smallest and that both our model and the single-regime GARCH model outperform the two regime-switching ARCH models. Apparently, in case of strong volatility persistence of individual shocks, taking account of this by GARCH terms is important. For the series with less volatility persistence of individual shocks, the dominance of our model over single-regime GARCH is larger, especially for the yen. Moreover, the forecasts generated by regime-switching GARCH and ARCH are of almost equal quality. So, using regimes as a source of volatility persistence is particularly important if the persistence of individual shocks is not very large.

The conclusions continue to hold when we use another, often-used forecast statistic, namely the coefficient of determination, R^2 , of the regression

$$(s\tau - \hat{\mu})^2 = \beta_0 + \beta_1 \hat{V}_{t-1}\{s\tau\} + \eta\tau. \quad (12)$$

This regression is comparable with the ones used by Pagan and Schwert (1990). The quality difference between the regime-switching models and the single-regime GARCH model, however, seems smaller now.

⁸The differences seem small, but relatively speaking they are substantial: the decrease in RMSE from the constant variance model to the single-regime GARCH model is enlarged by 6%, 21% and 117% for the pound, mark and yen, respectively, when going to the regime-switching GARCH model.

Using the R^2 as a forecast statistic has one drawback. It measures the quality of a linear combination, $\beta_0 + \beta_1 \widehat{V}_{t-1}\{s\tau\}$, of the forecast, although one is interested in the quality of the forecast itself. Therefore, we prefer the quality measure

$$R_{\beta=[0,1]'}^2 = 1 - \frac{V\{(s\tau - \widehat{\mu})^2 - \widehat{V}_{t-1}\{s\tau\}\}}{V\{(s\tau - \widehat{\mu})^2\}}, \quad (13)$$

which can be viewed as the R^2 under the restriction that $\beta_0 = 0$ and $\beta_1 = 1$. This forecast statistic is similar to the R^2 -type measure used by Gray (1996a). The tables make clear that this change from R^2 to $R_{\beta=[0,1]'}^2$ affects the quality measure for the single-regime GARCH model by far the most. The reason for this will become clear below. The seemingly smaller difference between the regime-switching models and the single-regime GARCH model thus appears to be entirely due to an incorrect quality assessment. Using $R_{\beta=[0,1]'}^2$ instead of the potentially misleading R^2 confirms our earlier conclusions based on the RMSE.

At first sight, it may seem that all models yield bad volatility forecasts, as the R^2 statistics are quite low. However, as Andersen and Bollerslev (1997) argued, it is naive to expect a "high" R^2 from regressions such as (12). They demonstrated that, even if $(s\tau - \widehat{\mu})^2$ is a conditionally unbiased estimator of the variance of interest, $V_{t-1}\{s\tau\}$, it is a very noisy one, which leads to a low R^2 . Using better proxies for the latent variable $V_{t-1}\{s\tau\}$, they showed that (single-regime) GARCH models do provide good volatility forecasts. Nevertheless, we find that regime-switching GARCH forecasts are better.

As stated before, the relatively poor forecasting performance of standard, single-regime GARCH models may well be caused by the high volatility persistence of individual shocks that we found in the previous subsection. Indeed, tables 5, 6 and 7 contain evidence for this claim. To explain this we analyze regression (12) again. If the mean and variance forecasts are (conditionally) unbiased, that is, $\widehat{\mu} = E_{t-1}\{s\tau\}$ and $\widehat{V}_{t-1}\{s\tau\} = V_{t-1}\{s\tau\}$, then regression (12) implies that $\beta_0 = 0$ and $\beta_1 = 1$ and that the error terms $\eta\tau$ are serially uncorrelated.

The first two implications of the unbiasedness of the forecasts, $\beta_0 = 0$ and $\beta_1 = 1$, are tested both individually and simultaneously using OLS estimates for β_0 and β_1 . For reasons of uniformity, all tests are corrected for autocorrelation and heteroskedasticity.⁹ The individual tests are robust t-tests, whereas the simultaneous test is a Wald test using a robust estimate of the asymptotic covariance matrix of the OLS estimator for $[\beta_0, \beta_1]'$.

Tables 5, 6 and 7 make clear that the hypotheses are overwhelmingly rejected for the

⁹In case of a ten-days forecast horizon the errors in (12) are no longer uncorrelated, so that the OLS standard errors have to be corrected; see also the notes below table 5.

single-regime GARCH model. All estimates for β_0 are significantly above zero, and all but one estimates for β_1 are significantly below one. Apparently, the GARCH variance estimates are too variable. Note that the rejection of $[\beta_0, \beta_1]' = [0, 1]'$ is exactly the reason behind the large difference between the often-used, but potentially misleading unrestricted R^2 and the better forecast statistic $R^2_{\beta=[0,1]'}$ defined in (13).

Regarding the estimates for β_0 and β_1 , all regime-switching models do much better, since only one coefficient is significantly different from its hypothetical value. Therefore, not surprisingly, the Wald tests for $\beta = [0, 1]'$ clearly prefer the regime-switching models over the single-regime GARCH model.

The crucial difference between the two types of models is that the regime-switching models have regimes as a means to capture volatility persistence, whereas single-regime GARCH models have to put all volatility persistence in the persistence of individual shocks. Given that the excessive variability of the forecasts disappears after the introduction of another way to capture volatility persistence, we conclude that it is the high persistence in single-regime GARCH models that makes the forecasts too variable.

To analyze the last implication of the unbiasedness of the forecasts, uncorrelated error terms $\eta\tau$ in (12), we compute the first-order autocorrelation in the residuals, ρ_1 , and the (heteroskedasticity consistent) modified Box-Pierce statistic \tilde{Q}_{10} introduced in table 1. The models that take most care of volatility clustering, standard GARCH, regime-switching ARCH(4;4) and GARCH, indeed show no significant autocorrelation. The highly significant values for ARCH(0;0) for the pound indicate once again that regime-switches alone are sometimes insufficient to capture all predictability of volatility.

Panel B of tables 5, 6 and 7 presents statistics for the **ten-days-ahead forecasts** ($\tau = t+9$). Most conclusions for the one-week-ahead forecasts also apply here. So, again the regime-switching GARCH model outperforms the single-regime GARCH model, especially for series with moderate persistence of individual shocks. The regime-switching GARCH model also outperforms the regime-switching ARCH models in case of strong volatility persistence within regimes. Overall, our regime-switching GARCH model is thus the preferred model for in-sample forecasting. Note that the outperformance of the four heteroskedasticity models over the constant variance model is lower than for the one-day forecast horizon, as all $R^2_{\beta=[0,1]'}$ are now closer to zero. Apparently, the longer the forecast horizon, the less valuable is the information in the information set I_{t-1} for forecasting. This is in line with the well-known fact that conditional heteroskedasticity is lower in lower-frequency data.

Out-of-sample forecasting

We now turn to the out-of-sample forecasts. We reestimate the five models for the three exchange rate returns using only the first half of the sample. Holding the parameters fixed, we then use the 2491 observations in the second half (from October 20, 1987 to July 12, 1997) to generate the volatility forecasts $\hat{V}_{t-1}\{s\tau\}$.¹⁰ As before, we take the one-day and ten-days horizons.

Tables 8, 9 and 10 show that the out-of-sample characteristics are similar to the in-sample ones. Single-regime GARCH models generate forecasts that are again too variable, which is not the case for the models that contain two variance regimes as a second way to capture volatility persistence. Secondly, the RMSE and $R^2_{\beta=[0,1]}$ demonstrate that the regime-switching GARCH model outperforms the standard, one-regime GARCH model, particularly in case of moderate volatility persistence of shocks, and that regime-switching GARCH again outperforms the two other regime-switching models for the British pound, which is the series with the highest persistence of shocks. The regime-switching forecasts for the other two series do not differ much, although the GARCH model is somewhat better for the German mark and ARCH(0;0) is slightly better for the Japanese yen.

7 Conclusion

Standard GARCH models often imply a lot of volatility persistence of individual shocks. We show that this makes the GARCH volatility forecasts too variable. Therefore, we extend the model such that shocks can, but need not be very persistent. This is achieved by the introduction of two regimes with different levels of volatility; a GARCH formula is used to specify the variance within a regime. In the resulting regime-switching GARCH model a shift from the high to the low variance regime can then explain that shocks are sometimes not persistent, but followed by a period of low instead of high volatility ("pressure relieving" effect). The empirical application using three U.S. dollar exchange rates shows the importance of regimes in a GARCH model: volatility forecasts improve substantially.

The "pressure relieving" effect, however, can also be explained by regime-switching ARCH models; why need a GARCH term in the regime-specific variance? Regime-switching ARCH models were developed, because it seemed to be practically impossible to allow for long-run dependence in the conditional variance caused by a GARCH

¹⁰We also did the reverse, that is, using the second half of the observations for estimation and the first half of the observations for forecasting. The conclusions are similar and will not be reported.

term. In this paper we present a way out of this problem by integrating out the unobserved regimes in a special way. Our empirical analysis demonstrates the usefulness of this: GARCH terms are important in regime-switching models and the quality of the volatility forecasts can benefit a lot from them.

Our third result is that volatility persistence is time-varying: in periods of high volatility shocks are more persistent than in periods of low volatility. Our model takes account of this by letting the ARCH and GARCH parameters differ across regimes. This time-varying volatility persistence reflects another way to decrease the effect of large shocks when necessary. A shift to the low variance regime after a large shock not only lowers the general volatility level, but it also implies a shift to the low persistence regime. Allowing for such effects appears to be empirically important.

As in other regime-switching models, the error distribution in our model is not restricted by normality; a t-distribution can be used as well. We show that this extra source of leptokurtosis is particularly useful in regime-switching models, since it makes the persistence of regimes less sensitive to outliers.

Another attractive feature of the model is that it yields simple, first-order recursive multi-period-ahead volatility forecasting formulas, similar to the ones for single-regime GARCH models. Moreover, estimation is also a first-order recursive process. Both properties make the model useful from a practical point of view.

Given the ease of generating volatility forecasts and that these forecasts outperform the often-used single-regime GARCH forecasts, the model has a number of possible applications. For example, the volatility forecasts can be used to analyze the effect of volatility on stock returns. Moreover, they may be helpful to price options, where volatility assessments are crucial. Both applications are left for future work.

Appendices

A Unconditional Error Variance

Suppose the "unconditional" variance $V\{\varepsilon_t | r_t\}$ exists for all $r_t = 1, 2$ and $\omega_1, \omega_2 > 0$. Using the variance definition (4), repeated use of the law of iterated expectations yields

$$\begin{aligned}
V\{\varepsilon_t | r_t\} &= \omega_{r_t} + \alpha_{r_t} E\{\varepsilon_{t-1}^2 | r_t\} + \beta_{r_t} E\{V_{t-2}\{\varepsilon_{t-1} | r_{t-1}\} | r_t\} \\
&= \omega_{r_t} + \alpha_{r_t} E\{E\{\varepsilon_{t-1}^2 | r_{t-1}, r_t\} | r_t\} + \beta_{r_t} E\{E\{V_{t-2}\{\varepsilon_{t-1} | r_{t-1}\} | r_{t-1}, r_t\} | r_t\} \\
&= \omega_{r_t} + \alpha_{r_t} E\{V\{\varepsilon_{t-1} | r_{t-1}\} | r_t\} + \beta_{r_t} E\{V\{\varepsilon_{t-1} | r_{t-1}\} | r_t\} \\
&= \omega_{r_t} + (\alpha_{r_t} + \beta_{r_t}) \cdot E\{V\{\varepsilon_{t-1} | r_{t-1}\} | r_t\}, \tag{14}
\end{aligned}$$

where the penultimate equality uses that the distribution of the error given the contemporaneous variance regime does not depend on the future variance regime r_t . Assuming that $\varepsilon_t | r_t$ is unconditionally homoskedastic, stacking $V\{\varepsilon_t | r_t = 1\}$ and $V\{\varepsilon_t | r_t = 2\}$ yields

$$\begin{bmatrix} V\{\varepsilon_t | r_t = 1\} \\ V\{\varepsilon_t | r_t = 2\} \end{bmatrix} = \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} + \begin{bmatrix} A_{1|1} & A_{2|1} \\ A_{1|2} & A_{2|2} \end{bmatrix} \cdot \begin{bmatrix} V\{\varepsilon_t | r_t = 1\} \\ V\{\varepsilon_t | r_t = 2\} \end{bmatrix}, \tag{15}$$

where $A_{i|j} = P\{r_{t-1}=i | r_t=j\}(\alpha_j + \beta_j)$. Expressions for the probabilities involved are given in appendix B. Let A be the matrix with elements $A_{i|j}$. Since we have assumed that both unconditional variances exist, $I_2 - A$ is invertible. We get the following one-to-one relation between the unconditional variances and the vector of ω_1 and ω_2 :

$$\begin{bmatrix} V\{\varepsilon_t | r_t = 1\} \\ V\{\varepsilon_t | r_t = 2\} \end{bmatrix} = (I_2 - A)^{-1} \cdot \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} \tag{16}$$

To obtain necessary conditions for the existence of both variances, we use $A_{2|1} = \alpha_1 + \beta_1 - A_{1|1}$ and $A_{1|2} = \alpha_2 + \beta_2 - A_{2|2}$ to rewrite $(I_2 - A)^{-1}$ as

$$(I_2 - A)^{-1} = \frac{1}{\det(I_2 - A)} \begin{bmatrix} 1 - A_{2|2} & \alpha_1 + \beta_1 - A_{1|1} \\ \alpha_2 + \beta_2 - A_{2|2} & 1 - A_{1|1} \end{bmatrix}. \tag{17}$$

Since the variances are strictly positive for all $\omega_1, \omega_2 > 0$, the four elements of $(I_2 - A)^{-1}$ must be nonnegative and $(I_2 - A)^{-1}$ may not have a zero row. Since $\alpha_i + \beta_i \geq A_{i|i}$

for both regimes $i = 1, 2$, this implies that $\det(I_2 - A) > 0$, so that $1 - A_{1|1} \geq 0$ and $1 - A_{2|2} \geq 0$. However, neither $A_{1|1}$ nor $A_{2|2}$ may be unity; otherwise $\alpha_i + \beta_i - A_{i|i} \geq 0$ for both regimes would imply that $\det(I_2 - A) < 0$, which is not the case.

In summary, we have three necessary conditions for the existence of the "unconditional" variances $V\{\varepsilon_t | r_t^v\}$, namely $A_{1|1}, A_{2|2} < 1$ and $\det(I_2 - A) > 0$. So, two linear combinations of the regime-specific ARCH and GARCH coefficients must be less than one and there is some restriction on the combination of the coefficients of both equations. These are similar to the necessary (and sufficient) condition $\alpha + \beta < 1$ in the standard, one-regime GARCH(1,1) model.

B Regime Probabilities

In section 3 we have to compute the density $p_{t-1}(s_t)$ in (7). This requires $p_{t-1}(r_t)$ and $p_{t-1}(r_{t-1} | r_t)$, where the latter is needed to compute the expectation in the variance $V_{t-1}\{\varepsilon_t | r_t\}$ given by (9). The first probability can be rewritten as

$$p_{t-1}(r_t) = \sum_{r_{t-1}=1,2} p_{t-1}(r_{t-1}) \cdot p_{t-1}(r_t | r_{t-1}), \quad (18)$$

where the constant switching probability $p_{t-1}(r_t | r_{t-1})$ follows from (6).

The second probability needed above is

$$p_{t-1}(r_{t-1} | r_t) = \frac{p_{t-1}(r_{t-1}) \cdot p_{t-1}(r_t | r_{t-1})}{p_{t-1}(r_t)}, \quad (19)$$

where $p_{t-1}(r_t)$ is given by (18).

Formulas (18) and (19) show that both regime probabilities in section 3 can be computed from one key probability $p_{t-1}(r_{t-1})$. This probability has a first-order recursive structure.

Recursion for Key Probability

The key probability can be written as

$$\begin{aligned} p_{t-1}(r_{t-1}) &= p_{t-2}(r_{t-1} | s_{t-1}) \\ &= \frac{p_{t-2}(s_{t-1} | r_{t-1}) \cdot p_{t-2}(r_{t-1})}{p_{t-2}(s_{t-1})} \end{aligned} \quad (20)$$

$$= \frac{p_{t-2}(s_{t-1} | r_{t-1}) \cdot \sum_{r_{t-2}=1,2} p_{t-2}(r_{t-2}) \cdot p_{t-2}(r_{t-1} | r_{t-2})}{p_{t-2}(s_{t-1})} \quad (21)$$

Because the switching probability $p_{t-2}(r_{t-1} | r_{t-2})$ is constant, the only variables needed to compute the basic probability are the previous key probability $p_{t-2}(r_{t-2})$ and the previous densities $p_{t-2}(s_{t-1} | r_{t-1})$ and $p_{t-2}(s_{t-1})$. This makes the computation of the key probability a first-order recursive process.

Regime Probabilities for Forecasting

Section 5 deals with forecasting l -periods-ahead exchange rate returns $s\tau$, where $\tau = t-1+l \geq t$. For their variance, $V_{t-1}\{s\tau\}$, we need in (10)

$$p_{t-1}(r\tau) = \sum_{r_{t-1}=1,2} p_{t-1}(r_{t-1}) \cdot p_{t-1}(r\tau | r_{t-1}), \quad (22)$$

where $p_{t-1}(r_{t-1})$ is the key probability discussed above. For the multi-period-ahead probability on the right-hand-side of (22), we form the time-constant Markov variance regime transition matrix M :

$$M = \begin{bmatrix} p_{11} & 1-p_{22} \\ 1-p_{11} & p_{22} \end{bmatrix}. \quad (23)$$

Using the l -th power of M , the theory of Markov processes states that

$$p_{t-1}(r\tau | r_{t-1}) = \left(M^l\right)_{r\tau r_{t-1}}. \quad (24)$$

To compute the variance $V_{t-1}\{\varepsilon\tau | r\tau\}$ in (11), where τ exceeds t , we need

$$p_{t-1}(r_{\tau-1} | r\tau) = \frac{p_{t-1}(r\tau | r_{\tau-1}) \cdot p_{t-1}(r_{\tau-1})}{p_{t-1}(r\tau)}. \quad (25)$$

The switching probability follows from (6) and the regime probability $p_{t-1}(r_{\tau-1})$ follows in a similar way as $p_{t-1}(r\tau)$ in (22); the denominator is given by (22).

Regime Probabilities in Unconditional Variance

In appendix A we need some unconditional probabilities to compute the unconditional error variance in (15). Using Bayes' rule, we have

$$p(r_{t-1} | r_t) = \frac{p(r_t | r_{t-1}) \cdot p(r_{t-1})}{\sum_{r_{t-1}=1,2} p(r_t | r_{t-1}) \cdot p(r_{t-1})}, \quad (26)$$

where $p(r_t | r_{t-1})$ is constant (see (6)) and the theory of Markov processes gives the unconditional probabilities (see Hamilton (1989)):

$$\begin{aligned} p(r_{t-1}=1) &= \frac{1-p_{22}}{2-p_{11}-p_{22}} \\ p(r_{t-1}=2) &= \frac{1-p_{11}}{2-p_{11}-p_{22}}. \end{aligned} \quad (27)$$

C Ex Post Regime Probabilities

As stated in section 4, the ex post regime probability $p_\tau(r_t)$, for a given future time $\tau \in \{t, t+1, \dots, T\}$, can be computed recursively, starting from the ex ante probability $p_{t-1}(r_t)$.

We can write $p_\tau(r_t)$ for both regimes $r_t=1, 2$ as

$$\begin{aligned} p_\tau(r_t) &= p_{\tau-1}(r_t|s_\tau) \\ &= \frac{p_{\tau-1}(s_\tau|r_t) \cdot p_{\tau-1}(r_t)}{\sum_{r_t=1,2} p_{\tau-1}(s_\tau|r_t) \cdot p_{\tau-1}(r_t)}, \end{aligned} \quad (28)$$

where $p_{\tau-1}(r_t)$ is known from the previous recursion for all $r_t=1, 2$, if $\tau > t$. If $\tau=t$, it is known from the estimation process, since then it is simply the ex ante probability given by (18).

The second ingredient of (28) is the density $p_{\tau-1}(s_\tau|r_t)$ for both regime outcomes. For $\tau=t$ it is known from the estimation process (see (8) and (9)), so that the filter probability, $p_t(r_t)$, follows directly from (28). Therefore, let us suppose from now on that $\tau > t$. Computing $p_{\tau-1}(s_\tau|r_t)$ requires a number of steps. We first write it as

$$p_{\tau-1}(s_\tau|r_t) = \sum_{r_\tau=1,2} p_{\tau-1}(s_\tau|r_\tau) \cdot p_{\tau-1}(r_\tau|r_t), \quad (29)$$

where we use that the conditional distribution of s_τ given r_τ does not depend on the earlier regime r_t . This formula itself has two ingredients. The first one is the density $p_{\tau-1}(s_\tau|r_\tau)$ for all regime combinations, which is known from the estimation process.

The second term needed in (29) is the $(\tau-t)$ -period-ahead regime-switching probability $p_{\tau-1}(r_\tau|r_t)$ for all regime outcomes. Once it has been computed, it should be saved, since it will be needed in the next recursive step. Making use of the Markov structure of the regime process, it can be written in terms of $(\tau-1-t)$ -period-ahead switching probabilities

$$p_{\tau-1}(r_\tau|r_t) = \sum_{r_{\tau-1}=1,2} p_{\tau-1}(r_\tau|r_{\tau-1}) \cdot p_{\tau-1}(r_{\tau-1}|r_t). \quad (30)$$

Again we have two ingredients. First, we need $p_{\tau-1}(r_\tau|r_{\tau-1})$ for all regime combinations. These are constant and follow from (6).

The second ingredient of (30) is $p_{\tau-1}(r_{\tau-1}|r_t)$ for all regime combinations. We have

$$\begin{aligned} p_{\tau-1}(r_{\tau-1}|r_t) &= p_{\tau-2}(r_{\tau-1}|r_t, s_{\tau-1}) \\ &= \frac{p_{\tau-2}(s_{\tau-1}|r_{\tau-1}) \cdot p_{\tau-2}(r_{\tau-1}|r_t)}{\sum_{r_{\tau-1}=1,2} p_{\tau-2}(s_{\tau-1}|r_{\tau-1}) \cdot p_{\tau-2}(r_{\tau-1}|r_t)}, \end{aligned} \quad (31)$$

where we use that the conditional density of $s_{\tau-1}$ is independent of the earlier regime r_t once $r_{\tau-1}$ is given. We have two ingredients. First, the conditional density $p_{\tau-2}(s_{\tau-1}|r_{\tau-1})$ for both regime outcomes. It is known from the estimation process. Secondly, we need the $(\tau-1-t)$ -period-ahead switching probability $p_{\tau-2}(r_{\tau-1}|r_t)$ for all regime combinations. This one was saved during the previous recursion, if $\tau > t+1$. If $\tau = t+1$, it equals one.

This completes the algorithm to compute (29), which is the second ingredient of (28). For each recursion one has to compute (31), use the result to compute (30) and use this to compute (29). Using this in (28) yields the ex post probability $p\tau(r_t)$ from $p_{\tau-1}(r_t)$. Therefore, starting from the ex ante probability $p_{t-1}(r_t)$ one can recursively compute the ex post probability $p\tau(r_t)$.

D Recursive Formula for Conditional Variance of Future Error

In this appendix we prove the recursive formula (11), which is needed for forecasting. It states

$$V_{t-1}\{\varepsilon\tau|r\tau\} = \omega_{r\tau} + (\alpha_{r\tau} + \beta_{r\tau})E_{t-1}\{V_{t-1}\{\varepsilon_{\tau-1}|r_{\tau-1}\}|r\tau\}, \quad (32)$$

where, for notational convenience, the index $t+i$ has been substituted by τ . This formula can be proved by repeatedly using the law of iterated expectations. Using definition (4), we get

$$\begin{aligned} V_{t-1}\{\varepsilon\tau|r\tau\} &= E_{t-1}[V_{\tau-1}\{\varepsilon\tau|r\tau\}|r\tau] \\ &= E_{t-1}\left[\omega_{r\tau} + \alpha_{r\tau}\varepsilon_{\tau-1}^2 + \beta_{r\tau}E_{\tau-1}\{V_{\tau-2}\{\varepsilon_{\tau-1}|r_{\tau-1}\}|r\tau\}|r\tau\right] \end{aligned} \quad (33)$$

For the ARCH part we get

$$\begin{aligned} E_{t-1}[\varepsilon_{\tau-1}^2|r\tau] &= E\{\varepsilon_{\tau-1}^2|r\tau, I_{t-1}\} \\ &= E[E\{\varepsilon_{\tau-1}^2|r_{\tau-1}, r\tau, I_{t-1}\}|r\tau, I_{t-1}] \\ &= E_{t-1}[E_{t-1}\{\varepsilon_{\tau-1}^2|r_{\tau-1}\}|r\tau], \end{aligned} \quad (34)$$

where the last equality uses that the error distribution given the contemporaneous variance regime does not depend on the future variance regime.

For the GARCH part we use similar techniques to obtain

$$\begin{aligned}
E_{t-1}[E_{\tau-1}(V_{\tau-2}\{\varepsilon_{\tau-1}|r_{\tau-1}\} | r\tau) | r\tau] \\
&= E[E(V\{\varepsilon_{\tau-1}|r_{\tau-1}, I_{\tau-2}\} | r\tau, I_{\tau-1}) | r\tau, I_{t-1}] \\
&= E(V\{\varepsilon_{\tau-1}|r_{\tau-1}, I_{\tau-2}\} | r\tau, I_{t-1}) \\
&= E[E(V\{\varepsilon_{\tau-1}|r_{\tau-1}, I_{\tau-2}\} | r\tau, r_{\tau-1}, I_{t-1}) | r\tau, I_{t-1}] \\
&= E[V\{\varepsilon_{\tau-1}|r_{\tau-1}, I_{t-1}\} | r\tau, I_{t-1}] \\
&= E_{t-1}[V_{t-1}\{\varepsilon_{\tau-1}|r_{\tau-1}\} | r\tau]. \tag{35}
\end{aligned}$$

The penultimate equality uses that $I_{\tau-2}$ given $r\tau, r_{\tau-1}, I_{t-1}$ is independent of $r\tau$, since the Markov structure implies that the distribution of variance regimes $\dots, r_{\tau-3}, r_{\tau-2}$ conditional on $r_{\tau-1}$ and $r\tau$ is independent of $r\tau$; this makes the returns $\dots, s_{\tau-3}, s_{\tau-2}$ also independent of $r\tau$ once $r_{\tau-1}$ and $r\tau$ are given.

Substituting the results for the ARCH and GARCH parts in (33) gives formula (32).

References

- Andersen T.G. and T. Bollerslev (1997), "Answering the Critics: Yes, ARCH Models Do Provide Good Volatility Forecasts," NBER Working Paper No. 6023.
- Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307-327.
- Bollerslev, T., R.Y. Chou and K.F. Kroner (1992), "ARCH Modeling in Finance," *Journal of Econometrics*, 52, 5-59.
- Cai, J. (1994), "A Markov Model of Unconditional Variance in ARCH," *Journal of Business and Economic Statistics*, 12, 309-316.
- Engle, R.F. (1982), "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation," *Econometrica*, 50, 987-1008.
- Engle, R.F. and T. Bollerslev (1986), "Modeling the Persistence of Conditional Variances," *Econometric Reviews*, 5, 1-50.
- Gray, S.F. (1996a), "Modeling the Conditional Distribution of Interest Rates as a Regime-Switching Process," *Journal of Financial Economics*, 42, 27-62.
- Gray, S.F. (1996b), "An Analysis of Conditional Regime-Switching Models," Working Paper, Fuqua School of Business, Duke University.
- Hakkio, C.S. and M. Rush (1989), "Market Efficiency and Cointegration: an Application to the Sterling and Deutschmark Exchange Markets," *Journal of International Money and Finance*, 8, 75-88.
- Hamilton, J.D. (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle", *Econometrica*, 57, 357-384.
- Hamilton, J.D. and R. Susmel (1994), "Autoregressive Conditional Heteroscedasticity and Changes in Regime," *Journal of Econometrics*, 64, 307-333.
- Lamoureux, C.G. and W.D. Lastrapes (1990), "Persistence in Variance, Structural Change, and the GARCH Model," *Journal of Business and Economic Statistics*, 8, 225-234.
- Newey, W.K. and K.D. West (1987), "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703-708.

- Newey, W.K. and K.D. West (1994), "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, 61, 631-653.
- Pagan, A.R. and G.W. Schwert (1990), "Alternative Models for Conditional Stock Volatility," *Journal of Econometrics*, 45, 267-290.
- West, K.D. and D. Cho (1995), "The Predictive Ability of Several Models of Exchange Rate Volatility," *Journal of Econometrics*, 69, 367-391.
- White, H. (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817-838.

Table 1: Moments of exchange rate returns and autocorrelation tests.

	British pound	German mark	Japanese yen
Mean	-0.00	0.00	0.01
Variance	0.43	0.47	0.46
Skewness	-0.04	0.04	0.44
Excess Kurtosis	3.04	2.70	3.78
Autocorr. ρ_1	0.07* (0.02)	0.02 (0.02)	0.01 (0.02)
Autocorr. \tilde{Q}_{10}	24.96* [0.01]	6.61 [0.76]	16.16 [0.10]
Autocorr. squares ρ_1^s	0.12* (0.01)	0.12* (0.01)	0.09* (0.01)
Autocorr. squares Q_{10}^s	539.20* [0.00]	371.24* [0.00]	164.37* [0.00]

Standard errors in parentheses and p-values in square brackets; * is significant at 5% level.

The first-order autocorrelation, ρ_1 , is estimated as the slope coefficient in a regression of the return, s_t , on the first lagged return, s_{t-1} , and a constant. The standard errors are based on White's (1980) heteroskedasticity-consistent asymptotic covariance matrix.

\tilde{Q}_{10} denotes a modified Box-Pierce type statistic that combines the first ten autocorrelations. Following Pagan and Schwert (1990), it is defined as the sum of the first ten squared normalized autocorrelation estimates, where the normalizing factors are the heteroskedasticity-consistent standard errors of the autocorrelation estimates. \tilde{Q}_{10} is asymptotically χ_{10}^2 distributed.

The first-order autocorrelation in the squared returns, ρ_1^s , and the Box-Pierce type statistic Q_{10}^s are similarly defined.

Table 2: Estimation results for the British pound.

		SINGLE REGIME		REGIME-SWITCHING		
		Constant variance	GARCH (1,1)	ARCH (0;0)	ARCH (4;4)	GARCH (2,0;1,1)
Mean	μ	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
Uncond. var. regime 1	σ_1^2	0.46* (0.02)	0.61* (0.18)	0.18* (0.01)	0.19* (0.02)	0.21 * (0.05)
ARCH regime 1	α_{11}		0.06* (0.01)		0.13* (0.05)	0.23 * (0.09)
	α_{21}				0.10* (0.04)	0.12 (0.06)
	α_{31}				0.02 (0.03)	
	α_{41}				0.02 (0.04)	
GARCH regime 1	β_1		0.93* (0.01)			
Uncond. var. regime 2	σ_2^2			0.64* (0.03)	0.56* (0.03)	0.52 * (0.07)
ARCH regime 2	α_{12}				0.07* (0.02)	0.05 * (0.01)
	α_{22}				0.04 (0.02)	
	α_{32}				0.06* (0.02)	
	α_{42}				0.06* (0.02)	
GARCH regime 2	β_2					0.94 * (0.01)
Degrees of freedom t-dist.	ν_1^{-1}	0.25* (0.02)	0.19* (0.01)	0.20* (0.03)	0.24* (0.03)	0.30 * (0.04)
	ν_2^{-1}			0.15* (0.02)	0.17* (0.02)	0.15 * (0.02)
Regime switch. prob.	p_{11}			0.984* (0.004)	0.992* (0.003)	0.989 * (0.005)
	p_{22}			0.986* (0.004)	0.996* (0.002)	0.997 * (0.002)
Log-likelihood minus GARCH(1,1)		-4681.60 -244.34	-4437.26 0	-4482.51 -45.25	-4454.37 -17.11	-4412.22 25.04
Autocorr. squared	ρ_1^s	0.11* (0.01)	0.01 (0.01)	0.04* (0.01)	-0.01 (0.01)	-0.00 (0.01)
normalized residuals	Q_{10}^s	533.80* [0.00]	5.30 [0.87]	53.92* [0.00]	30.26* [0.00]	9.14 [0.52]

Standard errors in parentheses and p-values in square brackets; * is significant at 5% level.

Both autocorrelation statistics have been defined below table 1.

For uniformity with other tables we do not report the first-order autoregressive coefficient that was estimated for the pound only.

Table 3: Estimation results for the German mark.

		SINGLE REGIME		REGIME-SWITCHING		
		Constant variance	GARCH (1,1)	ARCH (0;0)	ARCH (4;4)	GARCH (0,0;1,1)
Mean	μ	-0.00 (0.01)	-0.01 (0.01)	-0.00 (0.01)	-0.01 (0.01)	-0.00 (0.01)
Uncond. var. regime 1	σ_1^2	0.50* (0.02)	0.70* (0.20)	0.23* (0.02)	0.23* (0.02)	0.14 * (0.03)
ARCH regime 1	α_{11}		0.08* (0.01)		0.03 (0.03)	
	α_{21}				0.01 (0.02)	
	α_{31}				0.02 (0.03)	
	α_{41}				0	
GARCH regime 1	β_1		0.91* (0.01)			
Uncond. var. regime 2	σ_2^2			0.76* (0.04)	0.69* (0.05)	0.55 * (0.06)
ARCH regime 2	α_{12}				0.06* (0.03)	0.07 * (0.01)
	α_{22}				0.01 (0.02)	
	α_{32}				0.07* (0.03)	
	α_{42}				0.05* (0.02)	
GARCH regime 2	β_2					0.90 * (0.02)
Degrees of freedom t-dist.	ν_1^{-1}	0.25* (0.02)	0.20* (0.01)	0.17* (0.02)	0.19* (0.03)	0.26 * (0.06)
	ν_2^{-1}			0.14* (0.02)	0.15* (0.02)	0.18 * (0.02)
Regime switch. prob.	p_{11}			0.984* (0.004)	0.985* (0.004)	0.981 * (0.011)
	p_{22}			0.981* (0.005)	0.987* (0.004)	0.998 * (0.002)
Log-likelihood minus GARCH(1,1)		-4962.26 -183.92	-4778.34 0	-4802.59 -24.25	-4791.68 -13.34	-4768.97 9.37
Autocorr. squared	ρ_1^s	0.12* (0.01)	0.00 (0.01)	0.03* (0.01)	-0.00 (0.01)	0.00 (0.01)
normalized residuals	Q_{10}^s	371.42* [0.00]	8.03 [0.63]	28.46* [0.00]	27.00* [0.00]	9.41 [0.49]

Standard errors in parentheses and p-values in square brackets; * is significant at 5% level.

A 0 indicates a boundary solution.

Both autocorrelation statistics have been defined below table 1.

Table 4: Estimation results for the Japanese yen.

		SINGLE REGIME		REGIME-SWITCHING		
		Constant variance	GARCH (1,1)	ARCH (0;0)	ARCH (4;4)	GARCH (1,0;1,1)
Mean	μ	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)
Uncond. var. regime 1	σ_1^2	0.50* (0.03)	0.61* (0.11)	0.23* (0.02)	0.26* (0.03)	0.24 * (0.03)
ARCH regime 1	α_{11}		0.09* (0.02)		0.08* (0.04)	0.08 (0.05)
	α_{21}				0	
	α_{31}				0.05 (0.03)	
	α_{41}				0.02 (0.03)	
GARCH regime 1	β_1		0.87* (0.02)			
Uncond. var. regime 2	σ_2^2			0.71* (0.04)	0.68* (0.04)	0.64 * (0.05)
ARCH regime 2	α_{12}				0.04 (0.03)	0.06 * (0.02)
	α_{22}				0.05 (0.03)	
	α_{32}				0.05 (0.03)	
	α_{42}				0.01 (0.02)	
GARCH regime 2	β_2					0.78 * (0.10)
Degrees of freedom t-dist.	ν_1^{-1}	0.28* (0.02)	0.25* (0.02)	0.23* (0.03)	0.25* (0.03)	0.26 * (0.03)
	ν_2^{-1}			0.17* (0.02)	0.17* (0.02)	0.18 * (0.02)
Regime switch. prob.	p_{11}			0.976* (0.007)	0.982* (0.006)	0.977 * (0.008)
	p_{22}			0.975* (0.006)	0.982* (0.005)	0.983 * (0.005)
Log-likelihood minus GARCH(1,1)		-4794.20 -111.62	-4682.58 0	-4672.19 10.39	-4662.58 20.00	-4664.49 18.09
Autocorr. squared	ρ_1^s	0.09* (0.01)	0.02 (0.01)	0.04* (0.01)	0.01 (0.01)	0.01 (0.01)
normalized residuals	Q_{10}^s	164.36* [0.00]	13.02 [0.22]	21.57* [0.02]	11.11 [0.35]	11.74 [0.30]

Standard errors in parentheses and p-values in square brackets; * is significant at 5% level.

A 0 indicates a boundary solution.

Both autocorrelation statistics have been defined below table 1.

Table 5: In-sample volatility forecasting statistics for the British pound.

Forecast Statistic	SINGLE REGIME		REGIME-SWITCHING		
	Constant variance	GARCH (1,1)	ARCH (0;0)	ARCH (4;4)	GARCH (2,0;1,1)
Panel A: One-day horizon					
RMSE	0.951	0.919	0.931	0.925	0.917
R^2		0.065	0.039	0.049	0.067
$R^2_{\beta=[0,1]'}$		0.061	0.039	0.049	0.066
β_0		0.07* (0.03)	-0.05 (0.04)	-0.03 (0.04)	0.01 (0.04)
β_1	0.93 (0.06)	0.81* (0.08)	1.12 (0.13)	1.05 (0.11)	0.95 (0.10)
Wald for $\beta = [0, 1]'$	1.15 [0.28]	5.55 [0.06]	5.27 [0.07]	0.98 [0.61]	0.89 [0.64]
Autocorr. ρ_1	0.11* (0.02)	0.00 (0.02)	0.05* (0.02)	-0.02 (0.02)	-0.01 (0.02)
Autocorr. \tilde{Q}_{10}	157.61* [0.00]	2.89 [0.98]	33.03* [0.00]	14.66 [0.14]	3.00 [0.98]
Panel B: Ten-days horizon					
RMSE	0.956	0.938	0.943	0.944	0.932
$R^2_{\beta=[0,1]'}$		0.038	0.027	0.024	0.048
β_0		0.10* (0.03)	-0.10 (0.06)	-0.09 (0.05)	-0.00 (0.04)
β_1	0.93 (0.06)	0.72* (0.08)	1.23 (0.18)	1.19 (0.16)	0.96 (0.10)
Wald for $\beta = [0, 1]'$	1.16 [0.28]	12.56* [0.00]	4.51 [0.10]	5.47* [0.06]	0.98 [0.61]

Standard errors in parentheses and p-values in square brackets; * is significant at 5% level.

The R^2 , $R^2_{\beta=[0,1]'}$, β_0 , β_1 and the Wald test for $\beta = [0, 1]'$ all relate to regression (12).

The R^2 under the restriction $\beta = [0, 1]'$, denoted by $R^2_{\beta=[0,1]'}$, has been defined by (13).

The standard errors for the estimates of β_0 and β_1 in (12) and the asymptotic covariance matrix used in the Wald-statistic for $\beta = [0, 1]'$ have been corrected for autocorrelation and heteroskedasticity using the Newey and West (1987) asymptotic covariance matrix. Following West and Cho (1995), we have taken Bartlett weights and have used the same data-dependent automatic lag selection rule. This rule has certain asymptotic optimality properties and was introduced by Newey and West (1994).

The two heteroskedasticity-consistent autocorrelation statistics, ρ_1 and \tilde{Q}_{10} have been defined below table 1. They do not appear in panel B, since the unbiasedness of the forecasts no longer implies that the errors $\eta\tau$ in (12) are serially uncorrelated, as the forecast horizon now exceeds the one day period between observations (overlapping sample).

Table 6: In-sample volatility forecasting statistics for the German mark.

Forecast Statistic	SINGLE REGIME		REGIME-SWITCHING		
	Constant variance	GARCH (1,1)	ARCH (0;0)	ARCH (4;4)	GARCH (0,0;1,1)
Panel A: One-day horizon					
RMSE	1.023	1.004	1.001	1.000	1.000
R^2		0.044	0.041	0.043	0.045
$R^2_{\beta=[0,1]'}$		0.037	0.041	0.043	0.044
β_0		0.11* (0.03)	-0.06 (0.04)	-0.01 (0.03)	0.06 (0.03)
β_1	0.94 (0.05)	0.72* (0.06)	1.13 (0.11)	1.02 (0.09)	0.85 * (0.06)
Wald for $\beta = [0, 1]'$	1.49 [0.22]	24.64* [0.00]	4.62 [0.10]	0.50 [0.78]	5.24 [0.07]
Autocorr. ρ_1	0.12* (0.03)	0.03 (0.03)	0.05* (0.02)	0.01 (0.02)	0.02 (0.02)
Autocorr. \tilde{Q}_{10}	147.50* [0.00]	8.30 [0.60]	18.07 [0.05]	17.42 [0.07]	8.13 [0.62]
Panel B: Ten-days horizon					
RMSE	1.020	1.019	1.009	1.010	1.010
$R^2_{\beta=[0,1]'}$		0.003	0.020	0.019	0.019
β_0		0.19* (0.05)	-0.04 (0.07)	-0.06 (0.07)	0.09 (0.06)
β_1	0.93 (0.05)	0.52* (0.09)	1.08 (0.15)	1.10 (0.15)	0.76 * (0.12)
Wald for $\beta = [0, 1]'$	1.64 [0.20]	31.79* [0.00]	0.44 [0.80]	1.02 [0.60]	4.91 [0.09]

Standard errors in parentheses and p-values in square brackets; * is significant at 5% level.
For the definitions of the forecast statistics we refer to the notes below table 5.

Table 7: In-sample volatility forecasting statistics for the Japanese yen.

Forecast Statistic	SINGLE REGIME		REGIME-SWITCHING		
	Constant variance	GARCH (1,1)	ARCH (0;0)	ARCH (4;4)	GARCH (1,0;1,1)
Panel A: One-day horizon					
RMSE	1.099	1.093	1.087	1.085	1.086
R^2		0.023	0.023	0.027	0.026
$R^2_{\beta=[0,1]'}$		0.014	0.022	0.027	0.026
β_0		0.15* (0.03)	-0.05 (0.05)	-0.01 (0.05)	0.01 (0.04)
β_1	0.91 (0.05)	0.61* (0.06)	1.08 (0.13)	0.99 (0.13)	0.95 (0.09)
Wald for $\beta = [0, 1]'$	3.63 [0.06]	49.90* [0.00]	2.00 [0.37]	0.82 [0.66]	0.76 [0.68]
Autocorr. ρ_1	0.09* (0.04)	0.02 (0.04)	0.05 (0.03)	0.01 (0.03)	0.00 (0.03)
Autocorr. \tilde{Q}_{10}	59.44* [0.00]	12.05 [0.28]	11.88 [0.29]	5.84 [0.83]	6.93 [0.73]
Panel B: Ten-days horizon					
RMSE	1.086	1.092	1.081	1.082	1.082
$R^2_{\beta=[0,1]'}$		-0.006	0.009	0.008	0.008
β_0		0.24* (0.07)	-0.03 (0.08)	-0.01 (0.08)	0.02 (0.08)
β_1	0.91* (0.05)	0.40* (0.13)	1.04 (0.19)	0.98 (0.18)	0.92 (0.18)
Wald for $\beta = [0, 1]'$	4.14* [0.04]	35.66* [0.00]	0.46 [0.80]	0.67 [0.72]	0.87 [0.65]

Standard errors in parentheses and p-values in square brackets; * is significant at 5% level.

For the definitions of the forecast statistics we refer to the notes below table 5.

A negative $R^2_{\beta=[0,1]'}$ results when the variance of the difference between the squared forecast error and the variance forecast, $V\{(s\tau - \hat{\mu})^2 - \hat{V}_{t-1}\{s\tau\}\}$, is larger than that of the squared forecast errors only. This can happen when, for instance, the variance forecasts are very volatile.

Table 8: Out-of-sample volatility forecasting statistics for the British pound.

Forecast Statistic	SINGLE REGIME		REGIME-SWITCHING		
	Constant variance	GARCH (1,1)	ARCH (0;0)	ARCH (4;4)	GARCH (2,0;1,1)
Panel A: One-day horizon					
RMSE	0.865	0.835	0.842	0.841	0.834
$R^2_{\beta=[0,1]'}$		0.062	0.047	0.049	0.063
β_0		0.05 (0.03)	-0.05 (0.05)	-0.01 (0.05)	0.01 (0.04)
β_1	0.88 (0.07)	0.84* (0.08)	1.09 (0.14)	0.99 (0.13)	0.93 (0.10)
Wald for $\beta = [0, 1]'$	2.51 [0.11]	4.39 [0.11]	3.55 [0.17]	0.87 [0.65]	1.70 [0.43]
Autocorr. ρ_1	0.11* (0.02)	-0.01 (0.02)	0.03 (0.02)	-0.02 (0.02)	-0.00 (0.02)
Autocorr. \tilde{Q}_{10}	121.89* [0.00]	5.65 [0.84]	9.02 [0.53]	9.89 [0.45]	3.59 [0.96]
Panel B: Ten-days horizon					
RMSE	0.868	0.849	0.854	0.854	0.848
$R^2_{\beta=[0,1]'}$		0.043	0.031	0.031	0.046
β_0		0.06 (0.04)	-0.12 (0.07)	-0.18* (0.07)	-0.02 (0.05)
β_1	0.89 (0.07)	0.81* (0.09)	1.24 (0.18)	1.36 (0.20)	0.97 (0.11)
Wald for $\beta = [0, 1]'$	2.39 [0.12]	4.89 [0.09]	4.34 [0.11]	10.38* [0.01]	1.95 [0.38]

The first half of the sample has been used for estimation and the second half (2491 days) for forecasting. Standard errors in parentheses and p-values in square brackets; * is significant at 5% level. For the definitions of the forecast statistics we refer to the notes below table 5.

Table 9: Out-of-sample volatility forecasting statistics for the German mark.

Forecast Statistic	SINGLE REGIME		REGIME-SWITCHING		
	Constant variance	GARCH (1,1)	ARCH (0;0)	ARCH (4;4)	GARCH (0,0;1,1)
Panel A: One-day horizon					
RMSE	0.923	0.904	0.904	0.904	0.899
$R^2_{\beta=[0,1]'}$		0.040	0.039	0.040	0.049
β_0		0.12* (0.03)	0.03 (0.05)	0.04 (0.05)	0.06 (0.04)
β_1	0.89 (0.06)	0.68* (0.07)	0.88 (0.13)	0.85 (0.13)	0.81* (0.09)
Wald for $\beta = [0, 1]'$	3.04 [0.08]	19.14* [0.00]	1.73 [0.42]	3.22 [0.20]	5.17 [0.08]
Autocorr. ρ_1	0.14* (0.04)	0.03 (0.04)	0.06 (0.04)	0.06 (0.04)	0.02 (0.04)
Autocorr. \tilde{Q}_{10}	99.50* [0.00]	3.49 [0.97]	7.16 [0.71]	8.64 [0.57]	3.39 [0.97]
Panel B: Ten-days horizon					
RMSE	0.923	0.918	0.909	0.910	0.908
$R^2_{\beta=[0,1]'}$		0.017	0.028	0.027	0.033
β_0		0.14* (0.04)	-0.06 (0.08)	-0.09 (0.09)	0.01 (0.06)
β_1	0.89 (0.06)	0.59* (0.09)	1.07 (0.19)	1.09 (0.20)	0.89 (0.12)
Wald for $\beta = [0, 1]'$	3.04 [0.08]	25.51* [0.00]	2.21 [0.33]	5.13 [0.08]	4.17 [0.12]

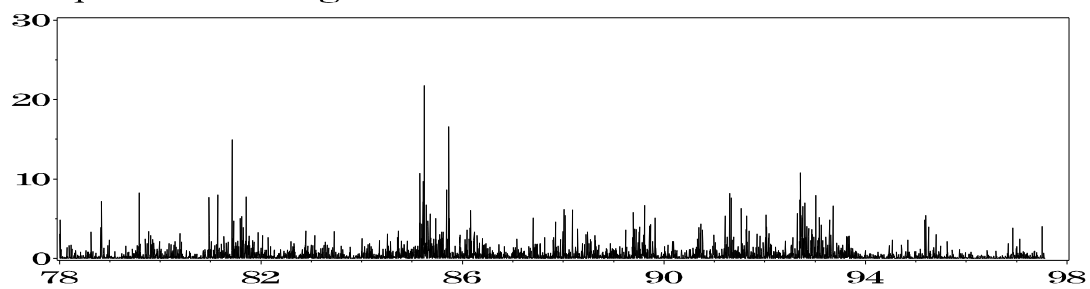
The first half of the sample is used for estimation and the second half (2491 days) for forecasting. Standard errors in parentheses and p-values in square brackets; * is significant at 5% level. For the definitions of the forecast statistics we refer to the notes below table 5.

Table 10: Out-of-sample volatility forecasting statistics for the Japanese yen.

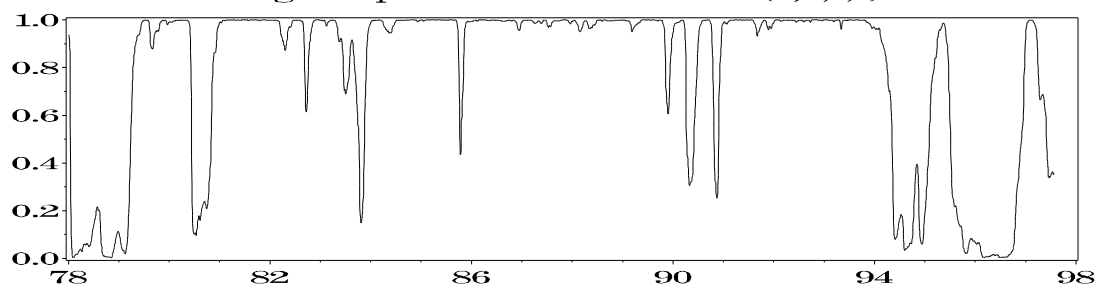
Forecast Statistic	SINGLE REGIME		REGIME-SWITCHING		
	Constant variance	GARCH (1,1)	ARCH (0;0)	ARCH (4;4)	GARCH (1,0;1,1)
Panel A: One-day horizon					
RMSE	1.001	0.998	0.987	0.991	0.991
$R^2_{\beta=[0,1]'}$		0.009	0.025	0.019	0.021
β_0		0.16* (0.04)	-0.14* (0.07)	0.06 (0.06)	0.04 (0.07)
β_1	0.87* (0.06)	0.55* (0.07)	1.26 (0.19)	0.80 (0.12)	0.83 (0.15)
Wald for $\beta = [0, 1]'$	4.18* [0.04]	45.86* [0.00]	9.63* [0.00]	3.91 [0.14]	4.32 [0.12]
Autocorr. ρ_1	0.09* (0.03)	-0.02 (0.03)	0.04 (0.03)	-0.00 (0.03)	-0.03 (0.03)
Autocorr. \tilde{Q}_{10}	69.45* [0.00]	8.12 [0.62]	11.43 [0.33]	12.31 [0.27]	11.24 [0.34]
Panel B: Ten-days horizon					
RMSE	1.001	1.010	0.995	0.998	0.998
$R^2_{\beta=[0,1]'}$		0.005	0.011	0.008	0.008
β_0		0.09 (0.09)	-0.34* (0.17)	-0.36 (0.20)	-0.44 (0.23)
β_1	0.87* (0.06)	0.58* (0.16)	1.66 (0.38)	1.61 (0.42)	1.75 (0.47)
Wald for $\beta = [0, 1]'$	4.18* [0.04]	37.56* [0.00]	6.72* [0.03]	9.18* [0.01]	10.73 * [0.00]

The first half of the sample is used for estimation and the second half (2491 days) for forecasting. Standard errors in parentheses and p-values in square brackets; * is significant at 5% level. For the definitions of the forecast statistics we refer to the notes below table 5.

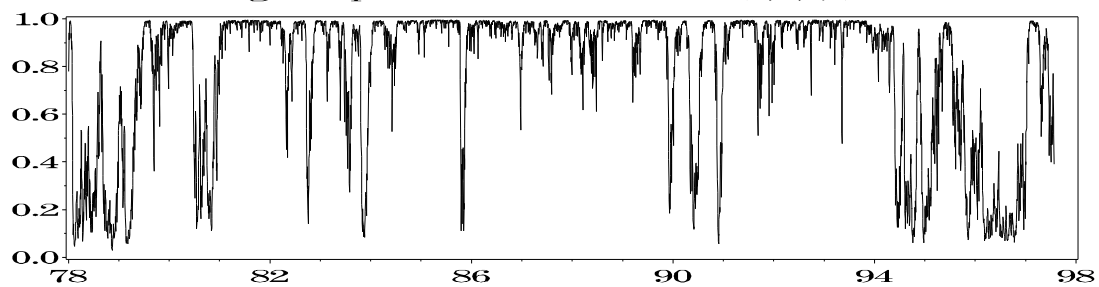
A: Squared exchange rate returns



B: Smoothed regime probabilities: GARCH(2,0;1,1)



E: Ex ante regime probabilities: GARCH(2,0;1,1)



F: Smoothed regime prob.: GARCH(2,0;1,1) under normality

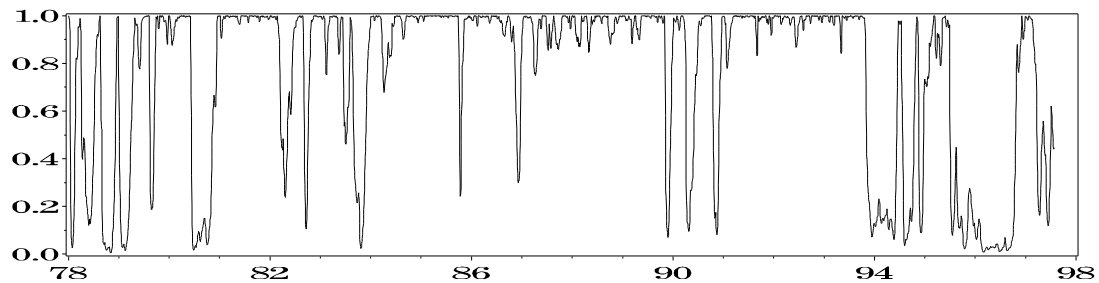
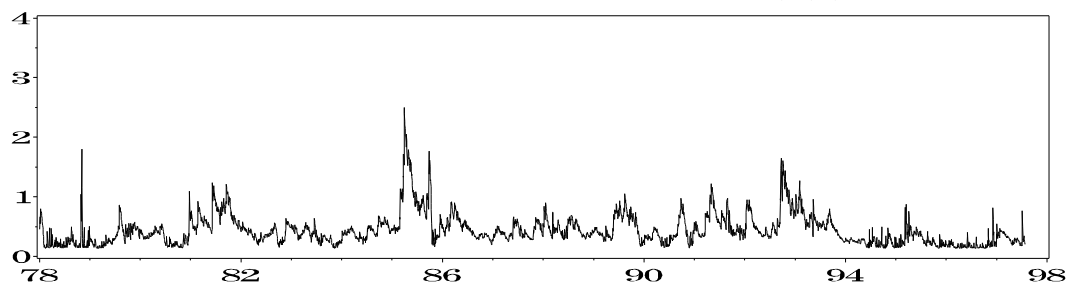
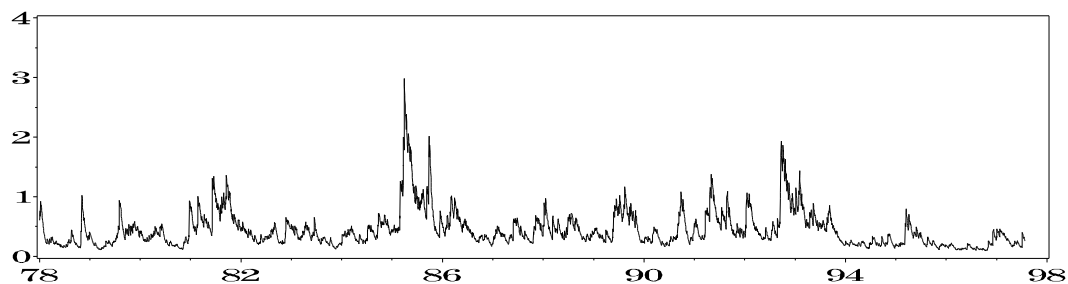


Figure 1: Continued on next page.

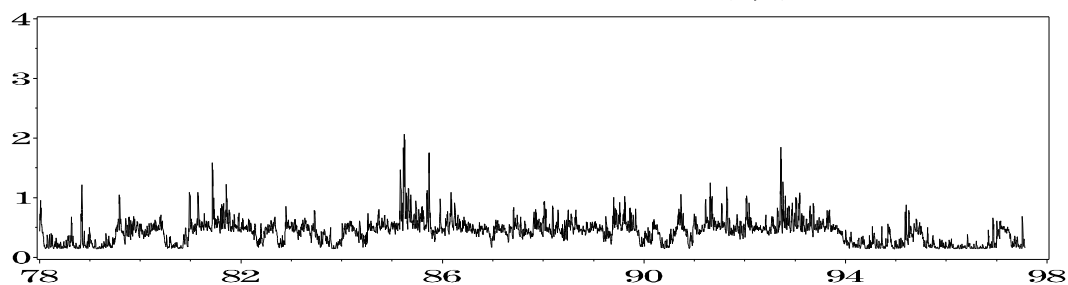
C: Conditional variance of returns: GARCH(2,0;1,1)



D: Conditional variance of returns: GARCH(1,1)



G: Conditional variance of returns: ARCH(4;4)



H: Cond. variance: GARCH(2,0;1,1) vs. GARCH(1,1)

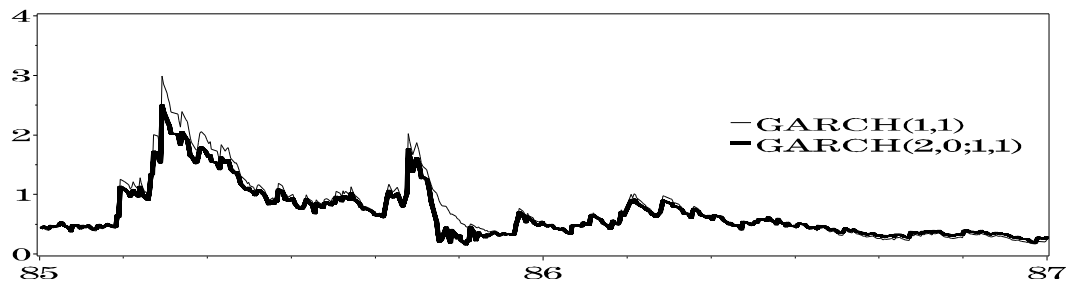
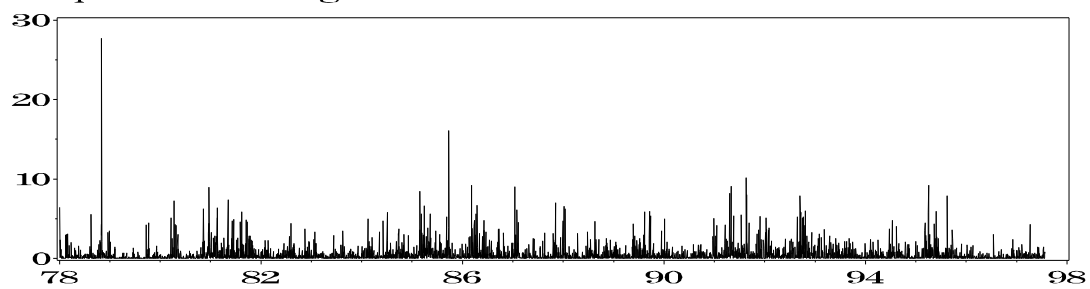
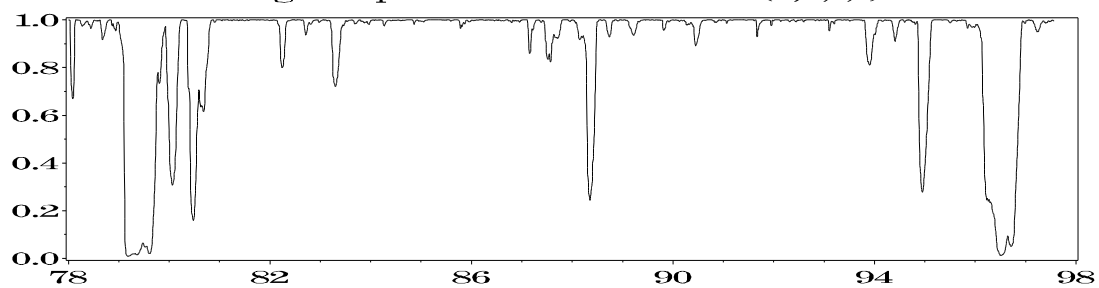


Figure 1: British pound over the sample period January 1978 to July 1997.

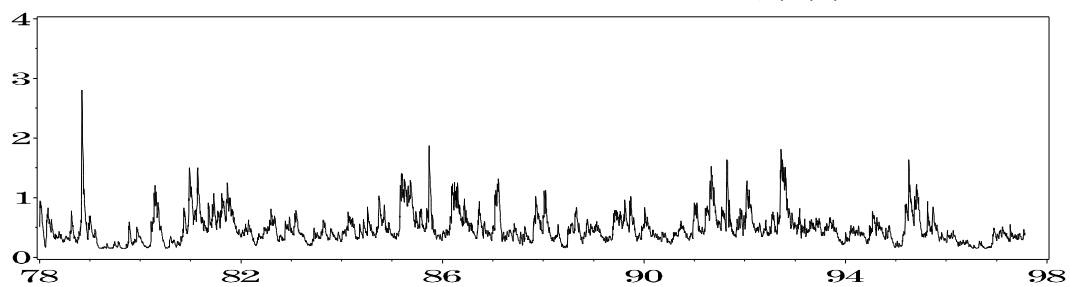
A: Squared exchange rate returns



B: Smoothed regime probabilities: GARCH(0,0;1,1)



C: Conditional variance of returns: GARCH(0,0;1,1)



D: Conditional variance of returns: GARCH(1,1)

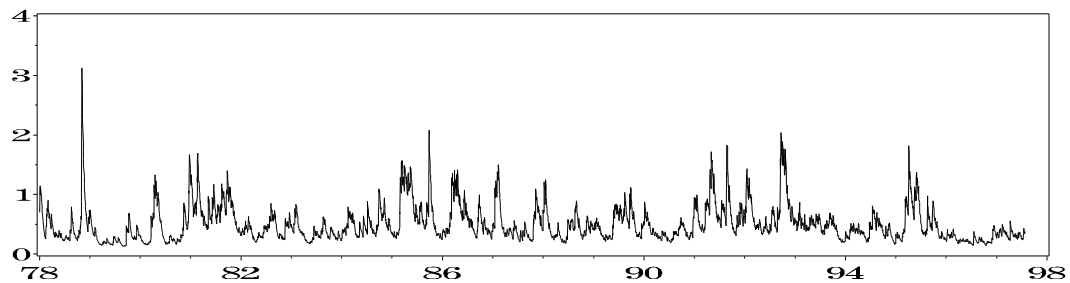
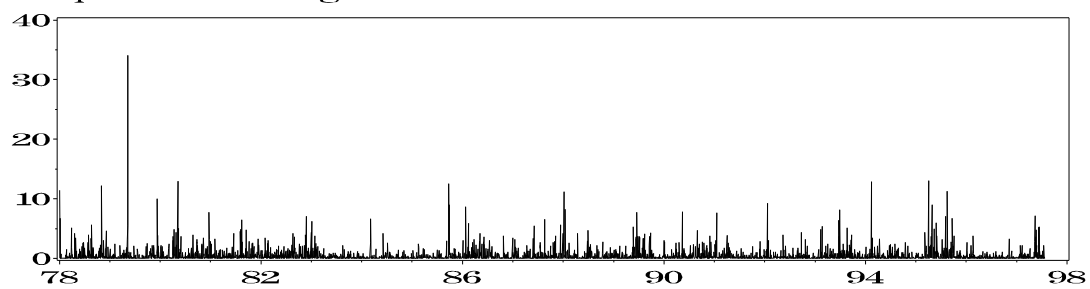
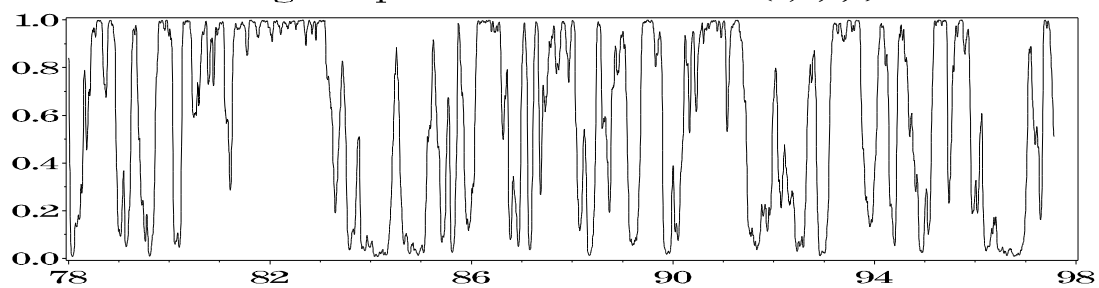


Figure 2: German mark over the sample period January 1978 to July 1997.

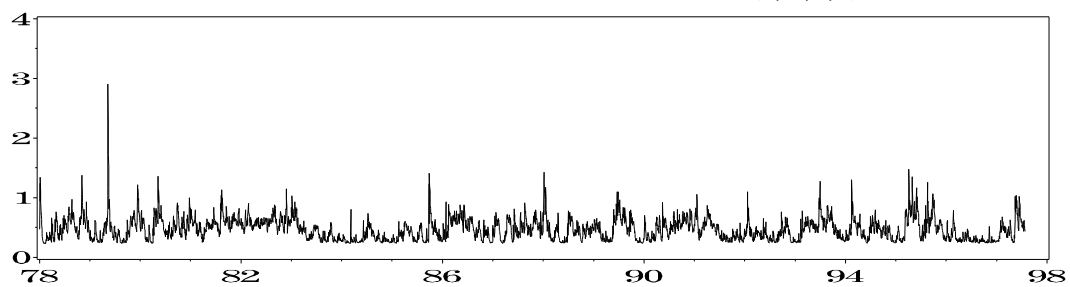
A: Squared exchange rate returns



B: Smoothed regime probabilities: GARCH(1,0;1,1)



C: Conditional variance of returns: GARCH(1,0;1,1)



D: Conditional variance of returns: GARCH(1,1)

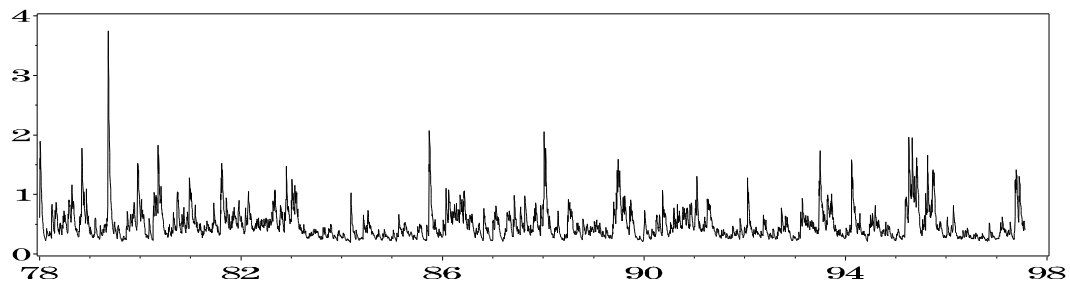


Figure 3: Japanese yen over the sample period January 1978 to July 1997.